DOCUMENT RESUME

ED 071 344 EM 010 489

AUTHOR	Hubert, Edwina E.; Rivers, LeRoy C.
TITLE	Introduction to Psychology and Leadership. An Analysis and Evaluation of Instructional Methodology for a Multimedia Course in Leadership, Psychology and Management.
INSTITUTION	Naval Academy, Annapolis, Md.; Westinghouse Learning Corp., Annapolis, Md.
SPONS AGENCY	National Center for Educational Research and Development (DHEW/OE), Washington, D.C.
REPORT NO	TR-6-11
BUREAU NO	BR-8-0448
PUB DATE	15 Oct 70
CONTRACT	N00600-68-C-1525
NOTE	178p.; See also EM 010 418 and EM 010 419
EDRS PRICE	MF-$0.65 HC-$6.58
DESCRIPTORS	Autoinstructional Aids; Communication (Thought Transfer); *Course Evaluation; Individualized Instruction; Individual Psychology; Instructional Design; *Leadership; Management Education; Multimedia Instruction; Programed Instruction; Program Evaluation; Social Psychology; Teaching Methods; Technical Reports

ABSTRACT

The effectiveness of a leadership training course (see the final reports which summarize the course development project, EM 010 418, EM 010 419, and EM 010 484) estimated from student performance on behavioral objectives is presented in this report. Descriptions and discussions of the course, of the development of materials and tests, of the media and presentation forms used, and of the procedures for course revision are provided. The instructional research involved in the project is also briefly covered. EM 010 420 through EM 010 447 and EM 010 451 through EM 010 512 are related documents. (Author/SH)

FILMED FROM BEST AVAILABLE COPY

Set #5

PR 6 043

ED 071344

Westinghouse Learning Corporation

AN ANALYSIS AND EVALUATION OF INSTRUCTIONAL METHODOLOGY FOR A MULTIMEDIA COURSE IN LEADERSHIP, PSYCHOLOGY AND MANAGEMENT

TR.6-11 October 15, 1970

Contract No. N00600-68-C-1525

EM 010 489

TR-6.11
October 15, 1970

ED 071344

AN ANALYSIS AND EVALUATION OF INSTRUCTIONAL METHODOLOGY FOR A MULTI-MEDIA COURSE IN LEADERSHIP, PSYCHOLOGY, AND MANAGEMENT

PHASE II EVALUATION REPORT

CONTRACT NUMBER N00600-68-C-1525

ABSTRACT

This report presents an assessment of the effectiveness of the first implementation of the United States Naval Academy multi-media course in Leadership, Psychology and Management. A description of the course, of the development of materials and tests, of the media and presentation forms used, and of the procedures for course revision is provided. A brief description of the instructional research involved in the project is also given.

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

Prepared by

Edwina E. Hubert
LeRoy C. Rivers

Approved by

Project Manager
Leadership Management Course

WESTINGHOUSE LEARNING CORPORATION

2083 WEST STREET

ANNAPOLIS, MARYLAND 21401

TABLE OF CONTENTS

SECTION I: Overview	1
SECTION II: Description of Course	4
1. Course Content and Objectives	4
2. Instructional Research	5
3. Media Selection	9
4. Course Organization	10
5. Implementation	14
6. Facilities	17
SECTION III: Description of Media and Presentation Forms	18
1. Videotapes	18
2. Audiotapes	19
3. Learning Activity Summary (LAS)	21
4. Linear Programed Text	22
5. Syndactic Text	24
6. Audiotape and Intrinsically Programed Text	27
7. Computer Assisted Instruction (CAI)	29
SECTION IV: Description of Tests	32
1. Test Item Pool	32
2. Administrative Pre and Posttests	33
3. Progress Checks	37
4. Cumulative Posttests	38

SECTION V: Results of Total Course Effectiveness 41

1. Total Course Effectiveness 44

2. Total Course Effectiveness Based on Administrative Tests 44

3. Total Course Performance Based on Progress Checks 54

SECTION VI: Part and Segment Effectiveness 61

1. Evaluation of Part Effectiveness 61

2. Evaluation of Segment Effectiveness 63

SECTION VII: Segments-Within-Media Effectiveness 70

SECTION VIII: Use of Effectiveness Data in Course Improvement 76

1. Materials Revision 76

2. Test Item Revision 87

APPENDICES

A. Course Outline 93

B. Guidelines for the Development of Behavioral Objectives 95

C. Outline of Course Structure and Media 100

D. Flowchart of the Student's Instructional Sequence 103

E. Example: Behavioral Objective and Criterion Referenced Test Item 104

F. Frequency Distribution of Item Difficulty and Discrimination: Administrative Posttest 105

G. Example: Item Discrimination Statistics 106

H. Test Item Statistics Summary: Administrative Posttest 107

I. Example: Item Response Analysis: Administrative Posttest 108

J. Frequency Distribution of Item Difficulty and Positive Discrimination Indices: Progress Checks 109

K. Frequency Distribution of Negative Item Discrimination Indices 110

L. Frequency Distribution of Administrative Pre and Posttest Scores 111

M. Gain Ratios For All Students Based on Administrative Pre and Posttests 112

N. Test Item-Objective Reference: Administrative Pretest 116

O. Test Item-Objective Reference: Administrative Posttest 117

P. Example: Student Track 118

Q. Total % Objectives Achieved By Each Student Before and After Remediation 119

R. Example: Test Item-Objective Reference: Progress Checks 122

S.	Module Questionnaire	123
T.	Part and Segment: Average % Progress Check Items Correct, Student Time, and % Student Interest	124
U.	Average Percent Progress Check Items Correct, Student Time and % Student Interest By Segment Within Media	131
V.	Percent Students Scoring 70% or 80% on Progress Checks Before and After Remediation By Segment Within Media	141
W.	Example: Syndactic Text Frame Analysis	147

TABLES

TABLE 4.1	Reliability Coefficients, Means and Standard Deviations for Administrative Tests	34
TABLE 4.2	Reliability Coefficients for Cumulative Posttests	39
TABLE 5.1	Mean Gain Ratios Based on A Administrative Pre and Posttests	47
TABLE 5.2	Frequency Distribution of Gain Ratios	49
TABLE 5.3	Frequency Distribution of Number of Objectives Achieved By Percentage of Students: Administrative Tests	52
TABLE 5.4	Frequency Distribution of % Gain of Number of Objectives From Pre to Posttest	53
TABLE 5.5	Frequency Distribution of % Total Course Objectives Achieved By Number of Students: Progress Checks	56
TABLE 5.6	Frequency Distribution of Number of Objectives Achieved By Percentage of Students: Progress Checks	60
TABLE 6.1	Unweighted Mean % Progress Check Items Correct, Student Time, and % Student Interest For All Parts	62
TABLE 6.2	Frequency Distribution of Average % Progress Check Scores For All Segments	64
TABLE 6.3	Frequency Distribution of Average Student Time For All Segments	66
TABLE 6.4	Frequency Distribution of % Student Interest For All Segments	68

TABLE 7.1	Unweighted Mean % Progress Check Items Correct, Student Time, and % Student Interest For All Media	71
TABLE 7.2	Unweighted Mean % Students Scoring 70% or 80% on Progress Checks Before and After Remediation: Segments-Within-Media	74
TABLE 8.1	Test Item-Objective Reference (Part Five-Segment Four)	80
TABLE 8.2	Test Item-Objective Reference (Part Six-Segment Four)	82
TABLE 8.3	Revision Decisions Based on Progress Check Data	85

I. INTRODUCTION

The purpose of the first implementation of the USNA Leadership Management Course was to provide empirical data upon which the course could be objectively revised. In addition to assessing the effectiveness of the course materials in teaching stated behavioral objectives, the effectiveness of various media and presentation forms and their relationship to various student characteristics was experimentally investigated.

The evaluation of total course effectiveness was based on objective data collected on student performance over a series of behavioral objectives. The relative effectiveness of media and presentation forms was assessed from objective data based on a series of controlled comparisons of media and presentation forms within the instructional sequence.

In this report, only course effectiveness estimated from student performance on behavioral objectives will be presented. The effectiveness of media and presentation forms and their relationship to student characteristics will be presented in the Report of Phase II Research Findings (TR-6.12a & b). Effectiveness of the final implementation system as well as the cost effectiveness of the system will be presented in the Phase III Final Report.

Methods used in evaluating course effectiveness have been drawn largely from Technical Recommendations from the U. S. Office of Education. For a single volume coverage of evaluation methods actually employed, as well as alternate methods, see Some Statistical Problems in the Evaluation of Instructional Programs (Brennan, 1970).

Evaluation methods used in the first validation were selected on the basis of their capabilitites for yielding meaningful data for course revision. For this reason, methods such as control comparison of the experimental course vs. an existing course or follow-up study of course graduates were not utilized. Since results of these studies do not indicate where or how to revise course materials, these methods are appropriately used only after materials are finalized.

A description of the instructional system being evaluated is given in Section II. Media and tests which were used are given in Sections III and IV, respectively. Evaluation methods which were used are described throughout the results sections of this report. (Sections V, VI, and VII). The use of evaluation data in determining the adoption-rejection of course components and in identifying course components requiring revision is given in Section VIII.

The course has been revised on the basis of the data given in this report and the Report of the Phase II Research Findings. The Leadership, Psychology and Management course is being implemented again this fall with a replication of the instructional research. The results of the second course implementation will be given in the Phase III Final Report.

II. DESCRIPTION OF COURSE

For the benefit of readers unfamiliar with the USNA multi-media Leadership course, a brief description of the first implementation of the course and the terminology used throughout is given in this section.

COURSE CONTENT AND OBJECTIVES

The content outline for the USNA Leadership Management course was developed jointly by WLC and the USNA Naval Science Department. Sources for the content outline were WLC and USNA subject matter experts, texts on managment and naval science, and excerpted materials used in the traditional USNA Leadership Management course. Reference sources were presented in bibliography form at the end of the content outline for each part of the course.

Having decided the nature of the content, the content was analyzed and rough drafts of behavioral objectives were developed. In the process of developing objectives, the actual sequencing and outlining of total course content was completed. Accordingly, the course was composed of 12 parts (chapters) subdivided into 59 segments (lessons). A complete listing of part titles and segment headings is included in Appendix A (Course Outline). The number of segments per part ranged from two in Parts Two, Nine and Ten, to ten in Part Five.

The sequencing of topics within each level of the course was generally determined by SME input and precisely determined by behavioral analysis. See Sequencing Rationale. Specific behavioral objectives were developed at two levels for each topic heading within a segment. Terminal objectives (high level problem solving objectives) were developed for each major heading. Enabling objectives (lower level objectives) were developed where necessary to ensure the learning of information requisite to the attainment of terminal objectives. In the actual writing of behavioral objectives, WLC drew heavily from the guidelines offered by Robert Mager in Preparing Instructional Objectives. To ensure that behavioral objectives were representative of the varying levels or types of learning which should occur in the total curriculum, WLC devised a classification system to guide the writers. The classification system represented a composite of systems suggested by Bloom's taxonomy (1956), Gagne's eight types of learning (1965), and Evans, Homme and Glaser's RULEG System (1962). The guidelines are given in Appendix B.

INSTRUCTIONAL RESEARCH

Concommitant with the identification and sequencing of course content and objectives, WLC developed a series of instructional research hypotheses which were incorporated

in the first implementation of the course. A detailed explanation of the instructional research design can be found in the Report of Phase II Research Findings, Part I: Conditions of Instruction (TR-6.12a). In this report only the basic dimensions of the research are given.

Research hypotheses were based on concepts presented in A Behavioral Approach to Instructional Design and Media Selection, in which a distinction is made between the importance of media and presentation forms (Tosti and Ball, 1968). The distinction is that a medium is only the mode of transmission of information whereas the truly important characteristics of instruction are the forms of presentation of the information within the medium. Any medium can be described with respect to its characteristic presentation dimensions and with respect to its capability for varying dimensions of presentation. The most basic dimensions of presentation which characterize a medium are the form and frequency of the stimulus, response, and management decisions. In matrix form these dimensions are:

DIMENSIONS OF PRESENTATION

	Form	Frequency
Stimulus	Stimulus representation	Duration
Response	Response demand form	Response demand frequency
Management	Management form	Management frequency

Stimulus representation is the form in which the stimulus is presented. It can be written, spoken, or pictorial.

Stimulus duration is the temporal stability of the stimulus. Duration can be transient - such as movies and lectures, or persistent - such as textbooks or other printed matter.

Response demand form is the type of response the subject is required to make. The form of the response demanded can be covert, written, or spoken.

Response demand frequency is the frequency with which the subject is required to respond. The frequency with which questions are asked within an instructional sequence can be high, medium, or low.

Management form is the type of remediation the subject is given. It can be multilevel - remediation by differing levels of response demand frequency within a single form; multiform - remediation by a different form of presentation; repetition - a simple repeat of the same presentation; or error diagnostic - remediation by branching according to specific incorrect responses.

Management frequency is the frequency with which the presentation is repeated or changed according to the need for remediation. The frequency of management can be high, medium, or low depending on response demand frequency.

In line with the assumption that presentation variables were the critical elements in instruction rather than the medium itself, WLC developed five hypotheses, four of which were based directly on variations in presentation rather than variations in media. The hypotheses were:

> Hypothesis 1: Given both simple and complex tasks, transient presentations with high response demand frequency (HRDF) will be more effective than transient presentations with low response demand frequency (LRDF). There will be no difference in student performance between covert or overt response demand (RD) presentations, or between videotape and audiotape-panelbook media.
>
> Hypothesis 2: Given both simple and complex learning tasks, persistent presentation with high RDF will be more effective than persistent presentation with low RDF. There will be no difference in student performance between three conditions of RD (overt-written, overt-spoken, and covert).
>
> Hypothesis 3: High RDF will be more effective than low RDF with either high, medium, or low management frequency.
>
> Hypothesis 4: If performance is below criterion level with low RDF presentation, remediation with high RDF will result in a significant improvement in performance.
>
> Hypothesis 5: The use of "peer-monitor" procedures will significantly improve student performance in an instructional sequence.

Results of experiments based on these hypotheses will be presented in the Report of Phase II Research Findings (TR-6.12a & b).

MEDIA SELECTION

Media selected for the first validation of the instructional system were based on three major considerations: research design requirements, multi-media requirements, and implementation requirements.

1. Media included in the instructional research design were selected on the basis of their capabilities for varying presentation according to requirements of the experimental hypotheses. For example, hypothesis 1 required transient stimulus duration in addition to media comparison; therefore, videotapes and audiotapes were selected. By using both audiotapes and videotapes in parallel modules over the same content, it was possible to control all irrelevant presentation variables and study the effects of transient presentation across two media as well as comparing the effectiveness of the two media.

2. In addition to selecting media for their research capabilities, media were selected which would provide diversification in the instructional sequence. Although it was not the intention to directly compare media throughout the course, i.e., Medium A vs. Medium B, or Medium C vs. Medium D, a number of media were included so that some statement could be made about student performance via those media.

3. The third major consideration in media selection was the project plan for implementing and maintaining the multi-media system beyond the design stage. Of primary concern were the costs for developing, implementing, revising and maintaining the system. For example, material for certain media could be programed during the design stage of the project but would be prohibitively expensive to implement or maintain (revise) after the termination of the contract.

Another concern was that media fit conveniently into an individually paced program of studies without creating undue logistical problems. In this regard, special attention was given to existing facilities at the USNA, such as computer consoles, media centers, etc., and to the project availability of those facilities.

Media included in the first validation were: videotaped and audiotaped lectures with panelbooks, syndactic text, linear programed text, computer assisted instruction, learning activity summaries (bibliographies), film and group discussion. Detailed identification of the above media is given in Section III.

COURSE ORGANIZATION

The course organization was basically derived from a subdivision of content objectives and administration of tests based on those subdivisions. Content subdivisions are the part and segment. Two course units created as

a result of the instructional research were the module and the cumulative posttest unit.

Content

A part is equivalent to a chapter in the course. It is the largest content division and can be viewed in terms of global content objectives.

A segment is equivalent to a single lesson of approximately 40-60 minutes within a part. It is a collection of learning objectives that are closely related by content and educational purpose. To form a segment, behavioral objectives similar in content were identified and sequenced. For example, in Part Two, Individual Behavior, objectives based on "motivation" were grouped into a single segment (Part Two, Segment VI). Likewise, objectives based on "conflict" were grouped into a single segment (Part Two, Segment VII). The segment served as the logistical unit in implementation for purposes of scheduling and assessment of progress through the course materials. Depending on its purpose, a segment was classified as core, depth core, or enrichment.

Core segments are required segments which include all of the information requisite to the attainment of behavioral objectives. They were presented sequentially and were designed for implementation in an individually paced instructional system.

Depth core segments are associated with one or more segments and are directed toward amplifying the learning objectives of those segments. Depth core segments included in the first validation were film, group discussions, and classroom lectures by the USNA instructor. Unlike core segments, depth core were scheduled by the instructor with respect to time and place.

Enrichment segments are related to but not essential to the mastery of terminal objectives. They were optional to students who desired more information than that presented in core segments.

Instructional Research

Module. A module is a particular instructional condition used to prepare and deliver materials for a segment, identified in terms of the categories of the Tosti and Ball (1969) model. Several parallel modules were prepared in each segment utilized for research purposes, representing variations specified by the experimental designs. The different modules of a segment are distinguishable from one another by differences in presentation design and/or media, although the content is the same.

Cumulative posttest unit. The cumulative posttest (CPT) unit is a group of three or four adjacent segments within a part. There are 13 CPT units involving 45 of the 59 segments of the course. The primary criteria for

grouping segments into CPT units were that the segments dealt with similar types of content and objectives, and that the instructional sequences relating to particular concepts which were initiated in the unit would also terminate in the same unit. All segments in a CPT unit were developed in the same medium and with the same variations in instructional conditions between modules.

The outline of the course structure and media used is given in Appendix C. Hypothesis 1 was tested in CPT units 1, 3, 7 and 9. Hypothesis 2 was tested in CPT units 4, 6 and 10. Hypothesis 3 was tested in CPT units 5 and 13. Hypothesis 4 was tested in CPT unit 11. Section III of this report provides a brief overview of the various modules used within each of the media. These controlled experiments and the variations in presentation design (modules) utilized in each are discussed in detail in the Report of Phase II Research Findings, Part 1: Conditions of Instruction (TR-6.12a).

Test Organization

Four different tests were used throughout the course. They were the administrative pretest and posttest, the progress check, the cumulative posttest (CPT), and the USNA examination.

The administrative pre and posttest was an 80 point criterion referenced test composed of items representatively sampled from the objective-test item pool. There was at

least one administrative test item for each segment of the course. The pretest was given at the beginning of the course, and the posttest was given as part of the final examination.

The progress check was a criterion referenced test of approximately ten items. It was taken at the end of each segment.

The cumulative posttest (CPT) was a norm referenced research test composed of positively discriminating content-related test items. Each CPT was composed of ten items for each segment in the unit. Cumulative posttests were given at the end of each experimental unit.

USNA examinations were a combination of criterion referenced test items selected from the objective-test item pool and items developed by the USNA on-site instructor. These were the only test results in the course which were used to determine the midshipmen's grades.

IMPLEMENTATION

The first implementation of the course was conducted in the Spring of 1970. Forty-four USNA midshipmen participated. The course was administered by the USNA on-site instructor, the WLC on-site instructor, the WLC systems analyst, and two data clerks. Specific procedures

used in implementing the course are given in the Instructor's Guide (TR-6.6).

The instructor's basic responsibilities were tutoring students needing remediation, leading group discussions, scheduling and administering depth core segments, scheduling and administering examinations, and determining grades.

The systems analyst developed and supervised the logistical procedures of the course. Data clerks controlled dissemination and collection of all core materials, remediation prescription forms, module questionnaires, progress checks, and cumulative posttests (CPTs). They also scored progress checks and CPTs and forwarded data to WLC's computer center.

Students were routed through the course according to procedures outlined in the Student Guide (TR-6.5). In brief, students worked through core segments of the course at their own speed. They were allowed to check out software materials and study them whenever and wherever they wished. All students were given identical material when they studied a non-research segment; i.e., they were instructed by the same form of presentation. For research segments, they studied by the particular module (form of presentation) to which they were assigned. Students were randomly assigned to modules at the

beginning of the course. Each student received his own routing schedule which included not only the sequence of segments he must study but also the schedule for remediation, research tests, and USNA examinations.

Students worked through non-research material by studying a segment, taking a progress check, remediating (if necessary), and then retaking the progress check. The requirement for remediation was based on failure to attain 80% of the objectives as measured by the progress check. If the student failed to meet the 80% criterion on his first try, he was given a remediation prescription form which directed him to specific points in the content which related to the objectives failed. If the student failed to meet the 80% criterion following remediation, he reported to the on-site instructor for tutoring.

Students worked through research segments in the same manner as non-research segments except that they did not remediate until after they had completed the entire research unit and taken the cumulative posttest. A flow chart of the student's path through materials, progress checks, CPTs, and remediation is given in Appendix D.

FACILITIES

For the implementation of the course, WLC was provided three classrooms at the Naval Academy. One room, which was designated as the administrative office, contained desks for the administrative staff and storage space for half of the course materials (including tapes, printed material, tests, forms, and computer cards). The administrative room was used as the site for administrative conferences, for student tutoring, and for distribution and collection of all material.

The second room was used as the principal instruction room. It contained 15 student carrels equipped with Ampex VTR's (4900), TV monitors and earphones, and Craig cassette recorders.

The third room, used as a multi-purpose room, had three (fully equipped) carrels to handle overflow from the instructional room. In addition, there were 30 student writing desks which were used during depth core lectures, films, group discussions, and testing.

III. DESCRIPTION OF MEDIA AND PRESENTATION FORMS

Media used in the first implementation of the USNA leadership course were audiotapes with panelbooks, videotapes, linear programed texts, syndactic texts, audiotapes with intrinsically programed texts, and computers (see Appendix C). A description of the media and each of the variations in presentation form (modules) within each medium is given in this section.

VIDEOTAPES

Videotapes were used to present core content in lecture format. The lecturer for all videotape segments was an active Naval officer with teaching background. With the aid of a teleprompter, the lecturer presented all material verbatum from a prepared script. All videotapes were prepared at a commercial television station. Commercial quality quad video tape recorders were used which provided broad editing capability unavailable in one inch VTR format. Two cameras were used in taping the lecture to allow for integration of a series of visuals (charts, photographs, drawings, etc.). Additionally, key points of content were superimposed on the screen during the program. Four different modules were prepared which involved variations in the response demand frequency (RDF) and the response demand (RD) dimensions of the presentation design.

Module 1: (High RDF-covert response demand)

Module 1 of the videotape lecture was characterized by high response demand frequency. At appropriate points in the lecture, the lecturer referred the students to a numbered question in the panelbook. Students read the question and recorded their answers in the panelbook. Sufficient time was allowed for students to respond so they did not have to turn off the VTR. The number of questions asked ranged from 15 to 22. Module 4 differed from Module 1 only in that an overt response was required of the student.

Module 2: (Low RDF-overt response demand)

The low response demand module of the videotape lecture was developed by simply editing out all but three of the lecturer's references to questions in the panelbook. All other elements of the videotape lecture remained the same. Module 3 differed from Module 2 only in that a covert response was required of the student.

AUDIOTAPES

Audiotapes with panelbooks were used to present core content in lecture format. The lecturer was a commercial radio announcer. The lecturer presented all material verbatum from a prepared script. In all audio-video research segments, the scripts for audio and video presentations were identical, as were the modules employed.

All audiotapes were developed in a commercial recording facility. In developing the audio presentations, standard recording tapes were used. For student use, tapes were transferred to C-60 and C-90 cassette cartridges.

All charts, photographs, drawings, etc., accompanying the audiotape lecture were presented in a panelbook. (It should be noted that for videotape modules, these charts, etc., were presented on the VTR.)

Module 1: (High RDF-covert response demand)

Module 1 of the audiotape lecture was characterized by high response demand frequency. At appropriate points in the lecture, the lecturer referred the students to a numbered question in the panelbook. Students read the questions and recorded their answers in the panelbook. Sufficient time was allowed for students to respond so they did not have to turn off the cassette recorder. The number of questions asked ranged from 15 to 22. Module 4 differed from Module 1 only in that an overt response was required of the student.

Module 2: (Low RDF-overt response demand)

The low response demand module of audiotape presentations was developed by simply editing out all but three of the lecturer's references to questions in the panelbook. All other elements of the audio lecture

remained constant. Module 3 differed from Module 2 only in that a covert response was demanded of the student.

LEARNING ACTIVITY SUMMARY (LAS)

A learning activity summary is similar to a traditional textbook or bibliography approach to learning. It is a technique very often used in college and graduate seminars to put the responsibility for structuring learning and achieving objectives on the student. A learning activity summary was composed of three parts: 1) an overview of the segment, 2) behavioral objectives for the segment, 3) a bibliography of source material that was related to each of the objectives. Students worked through an LAS segment by reading the overview and objectives and studying text materials which were related to each objective. Text materials were either select pages in published text books or supplemental handouts.

The student studied all text material until he felt he had mastered the objectives and could pass the progress check. If he did not achieve 80% of the objectives on the progress check, he remediated by re-studying the prescribed text material. In Module 1 in this medium, pairs of students studied together. In Module 2 students studied individually.

LINEAR PROGRAMED TEXT

Linear programed texts used in the first validation were developed by the RULEG and EGRUL methods of programing (Rule-example; example-rule). These are essentially programing methods of presenting a rule (definition, principle) and having the student identify an example of the rule (from 2, 3, or 4 choices), or presenting an example and having the student identify the rule or principle which is depicted in the example (Markle, 1964). Variations of the RULEG-EGRUL method which were used are EG-EG and RUL-RUL. (See Appendix B, Guidelines for Development of Behavioral Objectives.)

It is important to note that although confirmation of responses is ordinarily an important part of programed instruction, confirmations were deleted in the first implementation in order to obtain valid data on student frame responses. The first three modules of linear text were presented in standard format; i.e., there was a question for every information frame. This defined the high response demand frequency (HRDF) dimension. The only difference in presentation among these modules was in the form of response required of the student.

Module 1: (HRDF-written RD)

Students were instructed to respond to each frame by writing their selection (A, B, C, or D) on the frame answer sheets.

Module 2: (HRDF-spoken RD)

Students were instructed to respond to each frame by speaking their selection (A, B, C, or D) into a tape recorder.

Module 3: (HRDF-covert RD)

Students were instructed to read each frame question and think the answers to themselves.

Modules 4, 5, and 6 of the linear text covered the same material as the first three modules, but the frequency of response demand varied within the presentations. Instead of asking a question for every frame, questions were asked for every second or third frame. In "no-question" frames, examples or principles, which would be deleted when questions were deleted, were reworded in statement form; e.g., instead of asking, "Which of these situations best exemplified principle X?," the frame was followed by a statement, such as "An example of principle X is ....." Modules 4, 5, and 6 differed from each other in the form of response demand.

Module 4: (LRDF-written RD)

This was the same as Module 1 except for low response demand frequency (LRDF).

Module 5: (LRDF-spoken RD)

This was the same as Module 2 except for low response demand frequency.

Module 6: (LRDF-covert RD)

This was the same as Module 3 except for low response demand frequency.

SYNDACTIC TEXT

A syndactic text is essentially a series of linear programed frames each preceded by a brief but complete summary of the information presented in the frames. Students worked through the syndactic text by reading the first summary statement and taking a summary quiz of 5 to 8 questions. If the student answered all summary quiz questions correctly, he read the second summary, took summary quiz 2, etc.

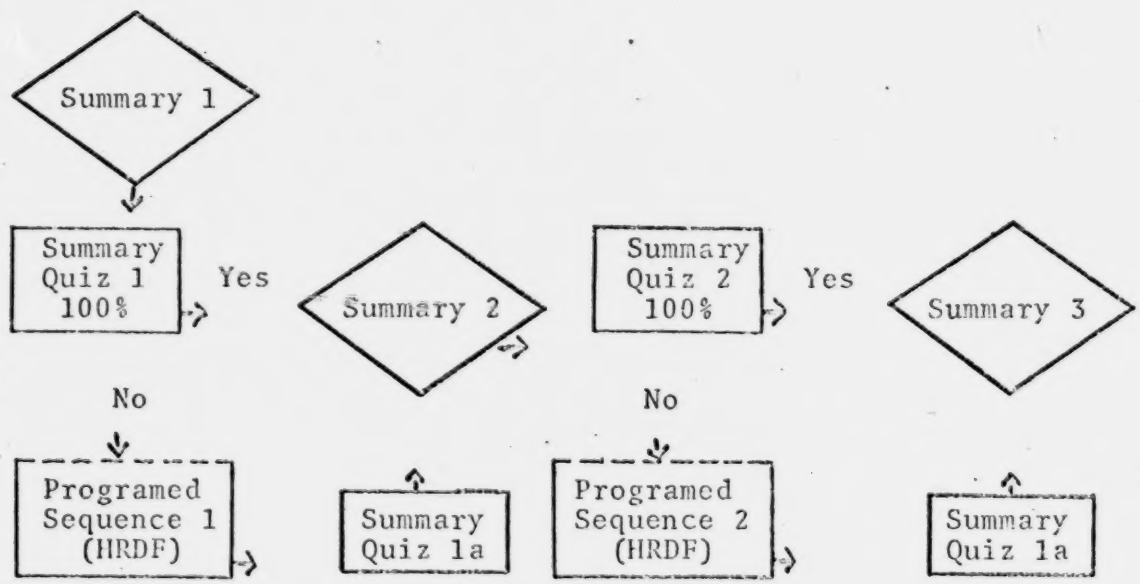
The student who incorrectly answered one or more questions of a summary quiz was required to study the linear programed sequence associated with that summary. The linear programed sequence was identical to the linear text discussed on page 21. It was developed by the RULEG or EGRUL method of presenting small bits of information accompanied by examples of the concepts being taught. At the end of the programed sequence, the student retook the summary quiz. Regardless of his performance when he retook the summary quiz, he went on to the next summary statement and repeated the procedure. Non-research segments of syndactic text were implemented according to the procedure given above.

Module 1: (HRDF)

The first module of research segments of syndactic texts was presented in the same manner as non-research segments. It was characterized by high response demand frequency(HRDF). Module 1 is depicted in chart form as:

PART TWO

SEGMENT SIX



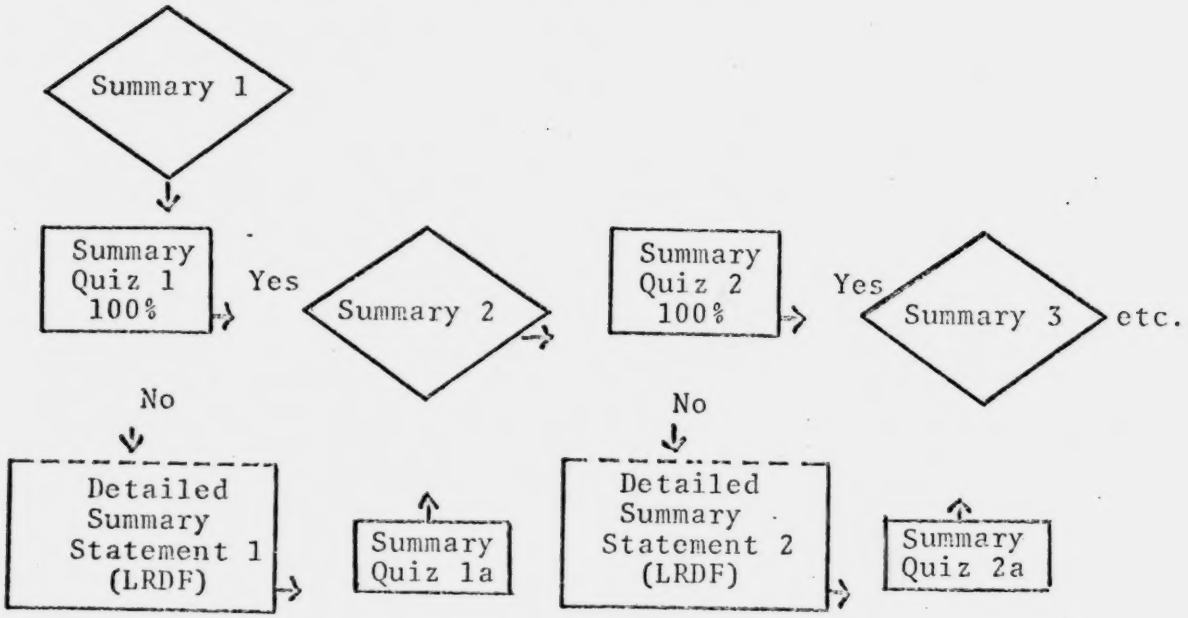
Module 2: (LRDF)

The second module of syndactic text was identical to Module 1 except that it was characterized by a low response demand (LRDF) program. Instead of having RULEG question frames, examples were simply given in statement form. This sequence of statements was referred to as a "detailed summary statement." There were no questions asked in the

detailed summary statement. Module 2 is depicted in chart form as:

PART TWO

SEGMENT SIX



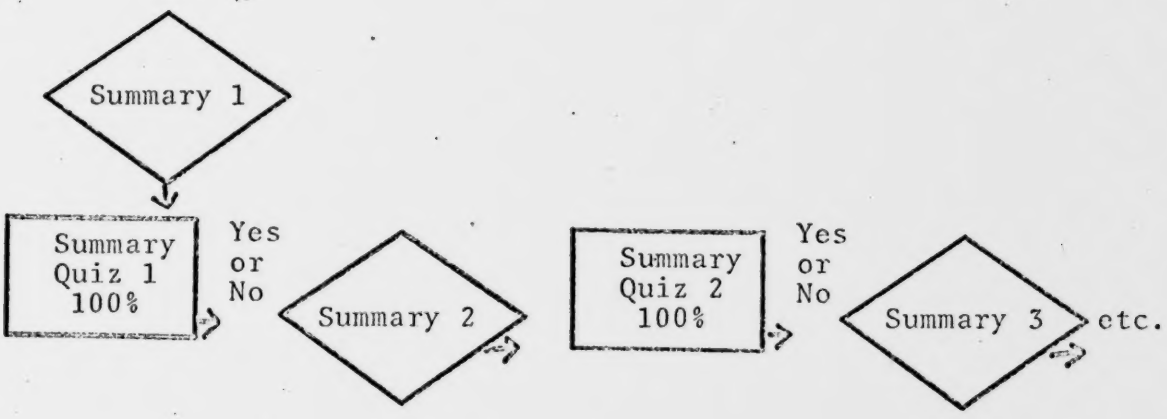
Module 3: (no RDF)

The third module was represented by summary statements alone (no remediation, no RDF). The student read a summary statement, took the summary quiz, and proceeded to the next summary statement regardless of his score on the summary quiz.

Module 3 is depicted in chart form as:

PART TWO

SEGMENT SIX



AUDIOTAPE AND INTRINSICALLY PROGRAMED TEXT

As originated by Norman A. Crowder, the intrinsic programing technique consisted of routing a student through a "scrambled" text on the basis of his response. Each response directed him to a different page of the text; thus, the student could not read through directly and sequentially.

Combining the intrinsic programing technique with an audiotape was a WLC innovation. In this teaching mode, the information was presented via the tape. While the student listened to it, he also looked at a summary page in the text which contained a precis of what he was hearing. He then stopped the tape and followed the instructions at

the bottom of the summary page, directing him to a page containing a question which tested the information given on the tape and summary. Each response to the test item referred the student to another page which informed him of the accuracy of that response. Thus, the student would select the alternative which he thought was correct, turn to the page indicated for that alternative, and find out if he had made a correct selection. If he had selected the correct response, he was instructed to go on to another summary page which he read while listening to the next audio portion. If his response was partially correct or incorrect, he was either told the nature of his error and instructed to proceed as described above, or he was instructed to return to the summary or question page to study the information again and select another alternative. This process of interaction between tape and text continued throughout the segment.

The tape, which contained the content of the segment, remained the same throughout the four modules. The text differed as follows:

Module 1: (HRDF-HMF)

In Module 1 a question was asked for each informational frame (HRDF). Based on his response, the student was always branched to a page where his answer was discussed and confirmed or rejected (high management frequency).

Module 2: (HRDF-MMF)

A question was asked for each frame (HRDF), but the student was branched on the basis of his response for only 50% of his responses (medium management frequency).

Module 3: (HRDF-LMF)

A question was asked for each frame (HRDF), but the student was never given feedback nor confirmed; i.e., he was never branched as a result of his selection. He simply went on to the next question (low management frequency).

Module 4: (LRDF-LMF)

Only three or four questions were asked throughout the programed sequence (LRDF), and the student was not branched on the basis of his responses to the three or four questions (low management frequency).

COMPUTER ASSISTED INSTRUCTION

Part Twelve (Applied Leadership) was developed as CAI material for the 1500 Instructional System. All of the three components of the system (CRT, audio, and image projector) were utilized in the implementation of the four modules. These modules paralleled the four modules used in the audiotape-IPB segments.

Module 1: (HRDF-HMF)

Module 1 was established in the following pattern: informational frame(s), questions, and feedback on the student's response. The informational frames were

presented on the CRT screen and image projector. The questions, which were often situations in which the student had to decide the best course of action, were presented: 1) on the audio, where the situation was described; 2) on the image projector, where pictures of the situation were presented along with the audio; 3) on the CRT screen, where the student was asked to select an answer from 3 to 5 choices. The student's selection, accompanied by feedback, was displayed on either the CRT screen or the audio tape, and occasionally on the image projector. This feedback consisted of the reason(s) why the selected answer was correct or incorrect.

Module 2: (HRDF-MMF)

Module 2 was developed from Module 1. The difference between the modules was in the amount of feedback. In Module 2, feedback was administered to a student's response only after every other question. In other words, the student was given a question, he responded, and he got feedback. He was then presented another question, he responded, and he got no feedback.

Module 3: (HRDF-LMF)

Module 3 was also based on Module 1. The student, however, got no feedback at all to any of his responses.

Module 4: (LRDF-LMF)

Module 4 contained the same material as Module 1. However, the student was presented only three questions in the entire segment. Instead of the questions that were presented in Modules 1 through 3, the material was presented in statement form. For example, a situation was described on the audiotape with a picture displayed on the image projector. Instead of asking the student a question pertaining to the situation, i.e., "What is the best way to handle the situation?," the student was presented a statement of the correct alternative to the question asked in the other modules, i.e., "The following is the best way to handle the situation." As in Module 3, no feedback was given.

IV. DESCRIPTION OF TESTS

The basic tests used to collect performance data during the first implementation were administrative pretest and posttest, progress check, and cumulative posttest. Brief descriptions of the development of each of these tests and the summary text statistics collected during their development are presented in this section. For detailed descriptions, see Research & Evaluation Plans, Part 1 (TR-6.3a).

TEST ITEM POOL

In order to ensure that at least the majority of course objectives were measured during the first validation, a test item pool of criterion referenced test items was developed. The original test item pool consisted of 1,416 criterion referenced items. The specification for developing test items was that two test items be developed for each of 12 objectives in a segment, i.e., 12x2x59=1,416. The selection of objectives to be included was based primarily on the need for representative coverage of terminal objectives, and secondarily on representative coverage of enabling objectives. All test items in the pool bore a one to one relationship to behavioral objectives and content. Since both WLC and USNA subject matter experts assisted in the development and review of the test items,

content validity for all items was assumed. (An example of a behavioral objective and its corresponding criterion referenced test item is given in Appendix E.) Objectives not covered by test items in the test item pool were measured within the course materials by criterion frames.

ADMINISTRATIVE PRE AND POSTTEST

Administrative tests were developed to provide a basis for evaluating total course achievement. The administrative pre and posttest was actually one 80-item test which was administered once at the beginning and once at the end of the course. (The administrative pre and posttest is distinct from USNA examinations which were administered throughout the course by the USNA instructor for purposes of assigning four-week and final course grades.)

The administrative pre and posttest was developed by representatively selecting test items from the test item pool. In this way, there was at least one test item selected from each segment of the course, plus an additional test item from each of 21 segments. Validity assumed for the administrative test was a high degree of content validity based on three SMEs' approval of test item-objective-content interrelationships. Reliability for administrative tests was estimated by the Kuder Richardson formula 20. Reliability coefficients, means, and standard

deviations for the administrative tests are given in Table 4.1.

TABLE 4.1

RELIABILITY COEFFICIENTS, MEANS AND STANDARD DEVIATIONS FOR ADMINISTRATIVE TESTS

	Mean	SD	r
Pretest	40.2	5.34	.52
Posttest	55.6	4.11	.41

It can be noted that the reliability for the pretest is greater than the reliability for the posttest. Perhaps the best explanation for the difference in pretest-posttest reliabilities is that while only 7% of the pretest item difficulties were 90% or higher, 27% or 36% of the item difficulties on the posttest were 90% or higher. The effect on reliability of having a test with many high difficulty items is the same as having a shorter test; with short tests, reduced reliability can usually be expected.

Within the present course, lack of posttest reliability is not considered to be an indication of poor test construction. Since administrative tests were composed of criterion referenced items, it was hoped that, in fact, all students would answer all items correctly, in which case the reliability of the posttest would be zero. In other words, the more successful the course and

the better the performance on each item of the posttest, the less measured reliability can be expected.

Frequency distributions of item difficulty and discrimination indices for the posttest are given in Appendix F. The item difficulty index is simply the percentage of students responding correctly to each item, i.e., the more student responses the higher the difficulty. The item discrimination index was computed by point bi-serial correlation of correct-incorrect responses and total test ranks for each item. Appendix G is an example of the type of calculations obtained for each item in computing the discrimination index.

In addition to the frequency distributions (Appendix F) and the individual item discrimination statistics (Appendix G), a concise summary of the item difficulty and discrimination index along with its significance for each item in the text was provided for the writer and analyst. This test item statistics summary for the posttest is given in Appendix H. Discrimination indices reported in Appendix H ranged from −.79 to .88. However, the wide range of discrimination is misleading because if the difficulty index for one item is either very high or very low, the discrimination index can be seriously affected by one high scoring student's incorrect response

or one low scoring student's correct response, i.e., the more uniform all students' responses to items, the more sensitive the discrimination index is to variations among the few remaining individuals.

It can be noted in Appendix H that although there is a wide range of discriminating items, only two discrimination indices were statistically significant. Since administrative tests are composed of criterion referenced items, it is not desirable to have many items which discriminate positively among students, and certainly undesirable to have items which discriminate negatively among students. In revising items for future implementations, more consideration was given to the selection of good growth items than to the selection of items which would discriminate among students. That is, items were selected on the basis of their ability to yield low difficulty on the pretest and high difficulty on the posttest.

In addition to item difficulty and item discrimination, an item response analysis was made of each administrative pretest and posttest question (see the example in Appendix I). From this analysis, the writers and analysts were able to determine the strength of item distractors as well as the general difficulty of the item. It was felt that the use of the ratings made by the students of the confidence

they had that the response they selected was correct would be of aid in revising the items. For example, if most of the students got an item correct but had very low confidence, the item would be reviewed for possible change. It was found, however, that most students responded with a rather high level of confidence in most situations. Therefore, confidence ratings associated with item responses were generally of limited value in revision.

PROGRESS CHECKS

Progress checks were developed to measure student achievement on each of the core segments. They were ten-item tests made up of criterion referenced items drawn from the test item pool. Ten items for each of the 59 core segments were drawn from among 24 items originally developed for each segment. Since they were drawn from the test item pool, which had met SME approval, content validity for progress checks was assumed. (No estimate of the reliability of progress checks is reported because the tests were composed of only ten criterion referenced items.)

Frequency distributions of item difficulty and item discrimination for all progress check items are given in Appendices J and K. It can be noted that most of the discrimination indices approach zero. High positive or negative discrimination indices are related

to very high or very low difficulty items (as is the case with administrative items). A general rule of thumb is: if the difficulty level of the item is below 40% or above 80%, the discrimination index computed by the item-total correlation method is questionable.

Analysis was made of all responses to each progress check item. An example of the format for item analysis is given in Appendix I.

CUMULATIVE POSTTESTS

Cumulative posttests were developed to assess the effects of variations in presentation across research segments. They were either 30- or 40-point tests designed to have high positive discrimination-low difficulty test items. CPT units are related to course content across segments in research units, rather than being directly related to terminal or enabling objectives.

Cumulative posttests are characterized by two types of test items: 1) simple associations, such as definitions and identifications, and 2) complex problem solutions, such as EGRULS. Validity for CPTs was assumed on the basis of representative sampling of content. There was considerable input to test item development by subject matter experts, although no attempt was made to have three independent judges review and signoff the test items. Reliability coefficients for CPTs were computed by Kuder Richardson-20 (Table 4.2).

TABLE 4.2

RELIABILITY COEFFICIENTS FOR CUMULATIVE POSTTESTS

Unit No.	No. of items	r
1	40	*
2	30	*
3	40	*
4	30	*
5	40	*
6	30	.64
7	40	.68
8	30	.58
9	40	.67
10	30	.18
11	30	.69
12	30	.65
13	40	.58

* Coefficients for these units were in error at the time of this reporting and are being re-computed.

As shown in Table 4.2, reliabilities for CPTs were reasonably high, considering the shortness of the tests. It is interesting to note that in contrast to administrative tests, high reliabilities for CPTs are desirable features of the tests. CPTs are designed for medium difficulty and positive discrimination among students and, as such, should show more item-item, item-total correlation than administrative tests. In revising CPT items, special attention was given to the discrimination level of CPT items as well as to item difficulty and item response distributions.

V. RESULTS OF TOTAL COURSE EFFECTIVENESS

Since the Leadership Management course is based on a behavioral approach to instruction, the first validation of total course effectiveness was based primarily on student test performance over a series of behavioral objectives. The premise for this form of evaluation is that once behavioral objectives are developed for a course and everyone agrees that they are necessary and worthwhile objectives, then the test of the effectiveness of the system is simply whether students attain the objectives.

The measurement of student performance on stated behavioral objectives is a technique for assessing the absolute effectiveness of a system. Relative effectiveness is assessed by comparing the effectiveness of one system or portions of the system to other systems. The relative effectiveness of the USNA multi-media system to one or more of the existing USNA leadership courses has not been assessed for the following reasons:

1. The evaluation of the multi-media course is based on student performance on over 600 behavioral objectives. A comparison of the effectiveness of the multi-media course to other courses would therefore necessitate the inclusion of over 600 measured objectives in the other courses. To compare effectiveness based on

final examinations alone would mean comparing effectiveness on only a sample of objectives rather than total objectives.

2. A second consideration in multi-media vs. existing leadership course comparisons is the possible Hawthorne and Rosenthal effects which may bias results. These two effects are, respectively, the tendencies 1) for students to realize they are in an experiment and to perform beyond typical expectations (Schramm, 1964), and 2) for teachers to realize they are being compared and thus alter their typical patterns of instruction (Rosenthal, 1966).

3. Assuming that effectiveness data for the multi-media course and the existing leadership course were not identical, there would be no way of accounting for the differences. Since several different media and forms of presentation are being used in the USNA course, and since teaching methods and materials vary from one USNA instructor to another, there would be no clear cut indication of the conditions of instruction which account for total differences.

4. Within the multi-media system, the effort is made to compare the relative effectiveness of one mode of presentation to another. Included in the forms of presentation employed are lectures presented via audiotape, textbooks, and bibliographies, all of which are traditional instructional techniques. In making these comparisons,

all variables except the relevant presentation variables are controlled; i.e., students are given the same content, objectives, materials, and test items. Only under these circumstances is it possible to state that one form of presentation is relatively more effective than another.

5. The multi-media course is not intended to supplant the instructor in other courses. It is intended rather to teach effectively the core content requisite to the understanding of Naval leadership, thereby reducing the need for the instructor's role as strictly a disseminator of information. An instructor using multi-media materials need only augment prepared materials with personal guidance of students. That is, he is able to lecture on points he would like to highlight, lead group discussions, tutor and counsel students, and in general use his time as a professional to invent new and creative ways of simulating leadership experiences. With these considerations in mind, studies of the effectiveness of systems which essentially compare one role of the instructor to another role are too global to be of value.

TOTAL COURSE EFFECTIVENESS

As mentioned in previous sections, the evaluation of course effectiveness has been based on student performance over two types of criterion referenced tests: the administrative pre and posttest and the progress check. Since both administrative and progress check tests consisted of items which had one-to-one correspondence with behavioral objectives, it is reasonable to estimate the average percentage of objective attainment by the average percentage of test items answered correctly. To the extent that test items and objectives have been directly matched and are thoroughly represented throughout the course, the percentage of objectives attainted and the percentage of test items achieved can be viewed interchangeably.

Results of total course effectiveness based on administrative tests represent estimates of effectiveness from a sample of objectives. Results of total course effectiveness based on progress checks represent estimates of effectiveness over all course objectives. Evaluation methodologies and results based on both types of tests are given in this section.

TOTAL COURSE EFFECTIVENESS BASED ON ADMINISTRATIVE TESTS

As discussed in Section IV, administrative pre and posttest was actually the same 80 point test given at the

beginning and end of the course. It consisted of items representatively sampled from each of the 59 core segments. It should be noted at this point that five items on the pretest were judged to be ambiguous by the WLC and USNA on-site instructors after the pretest had been administered. Since those five items were replaced for the posttest administration, both the five pretest items and the five replacements were dropped from the analysis of total test performance. Analysis was therefore based on 75 pre and posttest items rather than 80.

The most direct but somewhat misleading estimate of total course effectiveness is found in the average percentage of test items answered correctly on the administrative posttest, which is 74.1%. This estimate is misleading because: 1) it represents only the average percentage of objectives covered on the posttest (75 objectives), and 2) it does not take into account how much the students knew when they first entered the course (Deterline, 66th Yearbook NSSE, 1967). In light of this problem, to obtain a more comprehensive estimate of total course effectiveness, it is necessary to: 1) consider performance on all objectives measured throughout the course, and 2) consider the entering level of knowledge of the students. (Performance measures on all course objectives are given in Section VI.)

In order to assess the students' entering familiarity with the content, all students were pretested on the administrative test. The average percentage of pretest items achieved was 53.6%, and the average difference in amount achieved from pretest to posttest was 20.5% (74.1 - 53.6). (See Appendix L for the frequency distribution of pre and posttest scores.) The actual percentage gained from pretest to posttest is still a relatively obscure index of course effectiveness because it does not compare the amount of gain that was possible.

An index which does represent how much students learned from the system with respect to how much they could have learned is the ratio of actual gain to maximum possible gain. (Posttest minus pretest divided by the number of points on the test minus pretest.) The average crude gain ratio (raw score gain ratio) for the first implementation was 44.3% (Table 5.1). This indicates that, on the average, students learned approximately 44.3% of what they could have learned.

TABLE 5.1

MEAN GAIN RATIOS BASED ON ADMINISTRATIVE PRE AND POSTTEST

	Mean pretest	Mean posttest	Mean gain	Maximum possible gain	Gain ratio
Raw (crude) Score	40.2	55.6	15.4	34.8	44.3
True Score	40.1	55.5	15.4	34.9	44.4

N = 75

One problem associated with the use of the crude gain ratio is that it does not take into account the lack of reliability of either the pretest or posttest. Lack of test reliability has the effect of allowing varying results on the same test for the same students. For example, if a test were perfectly reliable, repeated testings of the same student would yield the same test scores. Proportionate to its lack of reliability, an unreliable test would produce varying test scores for the same student. The effect of test unreliability on the gain score would be that a student's score could vary unwards from his true score on

the pretest and vary downwards from his true score on the posttest, or vice versa. In either case, results based on the difference between pre and post testings of unreliable tests would yield unreliable data since it would indicate very small gain in the first case and large gain in the second.

In order to account for the lack of reliability of pre and posttests in computing the gain score ratio, estimates of the true gain score ratio for all students was computed (McNemar, 1958, and Horst, 1966).

Results in Table 5.1 indicate that the average estimated true gain and the true gain ratio, when corrected for test reliability, are virtually identical to the average crude gain and the crude gain ratio. However, it can be noted from Table 5.2 that the estimated true gain ratios for individual students are far more homogeneous than the crude gain ratios for individual students. (See also Appendix M for differences in true and crude pretest scores and true and crude gain for individual students.)

TABLE 5.2

FREQUENCY DISTRIBUTION OF GAIN RATIOS

Ratio	Crude gain frequency	Estimated true gain frequency
65.0-69.9	1	
60.0-64.9	3	
55.0-59.9	4	
50.0-54.9	8	
45.0-49.9	4	13
40.0-44.9	5	31
35.0-39.9	9	
30.0-34.9	5	
25.0-29.9	3	
20.0-24.9	2	
15.0-19.9		

A second problem associated with the use of gain score ratios in general is that there is no established standard by which to gauge them. To indicate that a system is 44% effective may seem to indicate that the system is not operating at an acceptable level of effectiveness. However, when the gain score ratio is interpreted in conjunction with the posttest average, it becomes more meaningful. Considering both the true gain score ratio and posttest average, it can be noted that students gained only 44% of what they could have gained, but their final scores were fairly high, i.e., 74.1%.

To further interpret the gain ratio and average posttest performance, it is necessary to consider each item of the test from which the data are derived. An additional element of data, therefore, is the test item-objective reference for the administrative pre and post-tests which indicates the number of objectives (test items) most students knew prior to taking the course and the number of objectives students failed to achieve on completing the course (Appendices N and O).

Refering to Table 5.3 (Frequency distribution of number of objectives achieved by % of students), it can be noted that there were no objectives which all students knew upon entering the course. However, by combining the first three frequency intervals, it can be seen that

80% or more students knew 17 of the items, or 23% of the 75 objectives on the pretest. In these cases, it is easy to understand that there was little room for gain on these items. On the other hand, although there were six items achieved by 100% of the students, by combining the last five frequency intervals for the posttest (Table 5.3), it can be noted that there were 11 items that were achieved by 50% or less of the students. This also accounts for low gain. From Table 5.4 it can be noted that there were additionally eight items which produced negative gain from pretest to posttest administrations; i.e., the students scored lower on the posttest than they did on the pretest. The use of the type of data discussed thus far in revising the course in outlined in Section VIII.

TABLE 5.3

FREQUENCY DISTRIBUTION OF NUMBER OF OBJECTIVES ACHIEVED BY PERCENTAGE OF STUDENTS: ADMINISTRATIVE TESTS

Percentage of students achieving objectives	Pretest number of objectives achieved	Posttest number of objectives achieved
100-	0	6
90-99	5	21
80-89	12	9
70-79	7	13
60-69	8	8
50-59	10	7
40-49	8	3
30-39	8	4
20-29	11	1
10-19	3	3
0-9	3	0

TABLE 5.4

FREQUENCY DISTRIBUTION OF PERCENTAGE OF GAIN OF NUMBER OF OBJECTIVES FROM PRE TO POSTTEST

Percentage of Gain	Number of Objectives
80 to 89	1
70 to 79	
60 to 69	5
50 to 59	3
40 to 49	7
30 to 39	7
20 to 29	10
10 to 19	15
0 to 9	19
-9 to 0	2
-19 to-10	4
-29 to-20	1
-39 to-30	1

Results from the data based on the administrative test item-objective reference (Appendix N and O) indicated two points quite clearly. One was that although students achieved an average of 74.1% on the posttest, a number of the test items were not adequate and needed to be revised. The second was that the objectives covered by the administrative test needed to be re-examined with respect to the adequacy with which they were taught. The percentage of correct student responses to items on the administrative tests is identical to the item difficulty level for these items. Procedures for revising the administrative exam based on the test item-objective reference are expanded upon in Section VIII.

TOTAL COURSE PERFORMANCE BASED ON PROGRESS CHECKS

As discussed in the preceding sections, estimates of total course effectiveness based on administrative tests represent student performance over only a sample of course objectives. Since administrative tests sampled only 75 of the course objectives, test results do not give a clear picture of student performance on each objective measured throughout the course which is needed to properly revise and improve the course.

In order to obtain data on all course objectives, progress checks were given at the end of each segment; they covered most of the terminal objectives and enabling objectives taught in those segments. (All enabling

objectives not covered by progress checks were covered by criterion frames within the program sequences.) Records of progress check results were made for each student, and the average progress check performance was computed for all students. The exact score on each progress check for each student was included in the student track. (An example of the student track is given in Appendix P.)

Since students took the progress check immediately after completing a core segment and then remediated on the basis of failure to attain 80% of the objectives, it is necessary to report progress check performance both before remediation (iteration 1) and after remediation (iteration 2). The percentage of total objectives attained by each student and the mean percent across students both before and after remediation are presented in Appendix Q. The frequency distribution of the percentage of total objectives attained by number of students (Table 5.5) is compiled from Appendix Q.

TABLE 5.5

FREQUENCY DISTRIBUTION OF PERCENTAGE OF TOTAL COURSE OBJECTIVES ACHIEVED BY NUMBER OF STUDENTS: PROGRESS CHECKS

% total objectives achieved	Number of students before remediation	Number of students after remediation
90.0-94.9		1
85.0-89.9		18
80.0-84.9	16	21
75.0-79.9	16	4
70.0-74.9	11	
65.0-69.9	1	

From Appendix Q it can be noted that the average percentage of objectives attained throughout the course prior to remediation was 77.7%. Following remediation, the average percentage of objectives attained increased to 83.8%. (These mean percentage figures were obtained by

$$\frac{\frac{a}{b}}{n}$$

where a = the number of objectives attained by all students, b = the number of progress check items given throughout the course (530), and n = the number of students in the course (44).)

As an indication of the variability of percentages of progress checks, the following groupings into percent-percent performance are reported:

> Before remediation, 97.8% or more of the students achieved 70% or more of the objectives.
>
> Following remediation, 91% or more of the students achieved 80% or more of the objectives.

These figures appear to indicate that, on the average, the course is performing quite well; however, the same problem in interpreting administrative posttest results is present in interpreting progress check results. That is, although 90% or more of the students are achieving 80% of the objectives, there are still some students performing below criterion level on some objectives. In

the first case, it can be noted from Table 5.5, that only four students failed to achieve the 80% criterion level after remediation, and their scores ranged from 76% to 79%. However, the second point (failure to achieve an average of 16% of the objectives after remediation) is more critical since this could represent up to 93 objectives throughout the course (16% of 530). Recall that the mean percent of objectives achieved after remediation (as shown in Appendix Q) was 83.8 which leaves roughly 16% of objectives not achieved.

Although this may seem high, it should be remembered that if a student did not achieve the objective after remediation, he received tutoring from the instructor. The revision process would be directed toward reducing the need for tutoring.

In order to determine if failure to attain 100% of the objectives was attributed to random errors on objectives or to certain poorly presented or poorly tested objectives, a test item-objective reference was developed for all progress check items (Appendix R). From the test item-objective reference, frequency distributions of the percentage of students responding correctly to each test item before and after remediation were developed.

By combining the first five frequency intervals in Table 5.6, it can be seen that 301 or 57% of the items were achieved by more than 80% of the students before remediation. Conversely, 43% of the objectives were not achieved by less than 20% of the students. These results indicated a need for revision of either 43% of the test items or corresponding materials or both. In order to localize whether materials or test items should be revised, the percentage of correct student responses to each test item following remediation was tablulated. Again by combining the first five frequency intervals, it can be noted from Table 5.6 that, following remediation, 404 or 76% of the test items were answered correctly by 80% or more of the students. Implications of these results for revision were that if students remediated through the same materials and were unable to answer the test item, the problem could still be either the test item or the materials. On the other hand, if they remediated through different materials and were able to answer the test item correctly, the item itself was probably adequate and the core materials should be revised. (Details of the use of progress check data in revision of materials is given in Section VIII.)

TABLE 5.6

FREQUENCY DISTRIBUTION OF NUMBER OF OBJECTIVES ACHIEVED BY PERCENTAGE OF STUDENTS: PROGRESS CHECKS

Percentage of students	No. of objectives achieved before remediation	No. of objectives achieved after remediation
100.0-	69	121
95.0-99.9	80	107
90.0-94.9	67	82
85.0-89.9	48	46
80.0-84.9	37	48
75.0-79.9	56	33
70.0-74.9	30	14
65.0-69.9	27	11
60.0-64.9	9	16
55.0-59.9	12	13
50.0-54.9	17	12
45.0-49.9	16	3
40.0-44.9	8	5
35.0-39.9	10	4
30.0-34.9	8	5
25.0-29.9	12	3
20.0-24.9	5	4
15.0-19.9	6	1
10.0-14.9	5	1
5.0- 9.9	5	1
0.0- 4.9	3	

VI. PART AND SEGMENT EFFECTIVENESS

Integral to the evaluation of total course effectiveness is the evaluation of independent parts and segments of the course. By assessing the effectiveness level of parts, and segments within parts, it is possible to identify and revise those segments which are not effective, thereby increasing total course effectiveness in future implementations.

EVALUATION OF PART EFFECTIVENESS

Evaluation of both parts and segments has been based primarily on student performance on progress checks before remediation. The total evaluation for parts is based on the average percentage of progress check items achieved, the average amount of time taken by students to complete each segment, and the percentage of students who indicated high or low interest in the materials before remediating. Summary results of analyses by part are given in Table 6.1. Student interest and time was tabulated from the module questionnaire shown in Appendix S. The students filled this out after each segment.

TABLE 6.1

UNWEIGHTED MEAN PERCENTAGE OF PROGRESS CHECK ITEMS CORRECT, STUDENT TIME, AND PERCENTAGE OF STUDENT INTEREST FOR ALL PARTS

Parts	Avg. % PC items correct	Avg. student time (min.)	% student interest (high)	% student interest (low)
Part One	74.4	60.9	56.4	8.6
Part Two	70.6	33.9	54.9	11.0
Part Three	73.9	30.8	36.0	19.4
Part Four	82.8	44.5	44.6	14.1
Part Five	73.3	45.8	22.4	23.8
Part Six	85.0	45.1	41.9	5.9
Part Seven	76.1	39.3	29.7	21.3
Part Eight	80.9	57.4	32.3	16.8
Part Nine	84.0	40.8	37.7	4.7
Part Ten	91.0	48.5	37.9	7.5
Part Eleven	84.6	31.5	31.4	6.1
Part Twelve	87.9	54.4	61.0	3.5

It should be noted that results by part are expressed in terms of the unweighted mean percentage of progress check items and student interest. This indicates that all data reported for a part represent the typical performance of segments within that part. For example, Part Two was characterized by segments in which 1) the typical performance on objectives was 70.6%; 2) the average amount of time (reported by students) per segment was 33.9 minutes; and 3) the typical reaction of students to segments was: 54.9% of the students reported the material as interesting, and only 11% reported it as uninteresting.

There are several reasons for variations among recorded data for parts. Some of the reasons are that data are averaged over differences in content, test items, media, and presentation forms. Differences can also be explained in terms of varying styles of writing among materials developers.

EVALUATION OF SEGMENT EFFECTIVENESS

Differences in student performance by segment are given in Appendix T. It should be noted that many of the same problems in interpreting differences between parts are also problems in interpreting differences between segments; i.e., segments cover different content, are tested by different progress check items, and are taught

by different media and forms of presentation. Even so, it can be observed from Table 6.2 that, for whatever reasons, performance on some segments is considerably better than others.

TABLE 6.2

FREQUENCY DISTRIBUTION OF AVERAGE PERCENTAGE OF PROGRESS CHECK SCORES FOR ALL SEGMENTS

Average percentage correct	Number of segments
90.0-94.9	7
85.0-89.9	9
80.0-84.9	13
75.0-79.9	11
70.0-74.9	7
65.0-69.9	10
60.0-64.9	1
55.0-59.9	1

It can be noted, for example, that in 29 of the segments students achieved an average of 80% or more of the objectives before remediation. In 12 of the segments, students achieved an average of less than 70% of the objectives. The data therefore indicate the need for substantial revision to either the test items or materials in 12 out of 59 or approximately 20% of the total core segments. (See Section VIII.) The test item-objective reference is used for identifying the test items, objectives, or content which require revision in these segments (Appendix R).

Table 6.3 presents the frequency distribution of average student time by segment. The average amount of time spent on each of the core segments ranged from 20 to 90 minutes.

TABLE 6.3

FREQUENCY DISTRIBUTION OF AVERAGE STUDENT TIME FOR ALL STUDENTS

Time in minutes	Number of Segments
90.0-94.9	1
85.0-89.9	
80.0-84.9	
75.0-79.9	
70.0-74.9	1
65.0-69.9	1
60.0-64.9	2
55.0-59.9	6
50.0-54.9	6
45.0-49.9	5
40.0-44.9	11
35.0-39.9	10
30.0-29.9	10
25.0-29.9	5
20.0-24.9	1
15.0-19.9	

It can be noted that although the range is 20 to 90 minutes, 93% of the segments had average times ranging from 25 to 65 minutes, or roughly from half an hour to an hour. Segments not falling within the half-hour to hour average time frame were revised with respect to projected estimates of student time required. It is important to remember, however, that the data reported on time are based on student self reports of the amount of time spent per segment prior to remediation. In other words, more time would be required for remediating if test performance on segments were not at criterion level. Therefore, more consideration was given to initial student test performance during revision than to time spent.

Student attitudes toward segment material were compiled from the module questionnaires. The first item on the questionnaire required the student to indicate on a five point rating scale whether his interest in the material was very much above average, above average, average, below average, or very much below average. The data were tabulated by collapsing the above average categories into percentage of high interest and collapsing the below average categories into percentage of low interest. The frequency distribution of percentages of students' high and low interest by segment are given in Table 6.4.

TABLE 6.4

FREQUENCY DISTRIBUTION OF PERCENTAGE OF STUDENT INTEREST FOR ALL SEGMENTS

Percentage of student interest	High frequency	Low frequency
80.0-84.9	1	
75.0-79.9		
70.0-74.9	2	
65.0-69.9	2	
60.0-64.9	5	
55.0-59.9	3	
50.0-54.9	3	
45.0-49.9	1	
40.0-44.9	9	
35.0-39.9	6	3
30.0-34.9	9	2
25.0-29.9	8	3
20.0-24.9	5	10
15.0-19.9	3	8
10.0-14.9		8
5.0- 9.9	1	12
0.0- 4.9	1	13

It can be noted that many more segments were judged to be of high interest than of low interest. On the other hand, there were 18 segments in which 20% - 40% of the students indicated low interest. Interpretation of segments with either high or low interest should be tempered with the realization that students' attitudes toward material may represent attitudes toward the content, toward the medium, toward the form of presentation or toward the quality of the presentation of the material. It can be noted from Appendix T that students generally reacted favorably to segments in which they also performed well. Whether students performed well because they liked the material, or whether they liked the material because they performed well is not clear. In either case, the revision of materials was guided primarily by students' performance on objectives, and secondarily by the average student time and student interest.

VII. SEGMENTS-WITHIN-MEDIA EFFECTIVENESS

As discussed in the previous section, there are problems with the interpretation of differences in part effectiveness because results may reflect differences in content, test items, media, or presentation forms. In an attempt to localize at least one of the variables which may account for differential effectiveness, segments taught by the same media have been grouped and analyzed.

At the outset, it is important to mention that although segments have been grouped on the basis of media, the results indicated in this section should not be construed as evidence for the superiority or inferiority of one medium to another. These results do not reflect inherent qualities of the media as such, but are rather indications of the effectiveness of the materials which were developed for and presented by each medium. For example, it could be that videotapes as a mode of transmitting X type of information in course Y is the best possible medium which could be selected; however, according to the way they are used in the multi-media course, video-tapes may be less effective than other forms of presentation. The reason for grouping and reporting

segment results by media is to localize the variations in effectiveness of segments which may be attributable to teaching via different media or forms of presentation. Results reported in this section are simply the average results obtained for all segments taught by a single medium throughout the course. The results do not indicate comparisons of media made over identical content with identical test items. Results of segment effectiveness by media are given in Table 7.1.

TABLE 7.1

UNWEIGHTED MEAN % PROGRESS CHECK ITEMS CORRECT, STUDENT TIME, AND % STUDENT INTEREST FOR ALL MEDIA

Media	% Progress Check items	Average student time (min.)	% Student interest (high)	% Student interest (low)
CAI	87.9	54.4	61.0	3.5
LAS	73.8	57.1	30.3	21.5
Syndactic text	80.2	43.2	47.5	10.6
Videotape	73.5	43.2	31.1	17.3
Audiotape	74.8	43.8	30.7	21.3
Audiotape IP	84.4	47.7	44.7	10.2
Linear Text	78.7	48.5	32.1	21.1

It can be noted that, on the average, the most to the least effective segments were those presented by: 1) CAI, 2) audiotape-IP booklets, 3) syndactic text, 4) linear texts, 5) audiotape lectures, 6) learning activities summaries, and 7) videotapes. Although there are not large differences between audiotape, videotape, and LAS segments, there are larger differences between these segments and CAI segments.

Correspondingly, it can be noted that student interest reports favored segments-within-media in approximately the same rank order as student performance; i.e., CAI, audiotape-IP, and syndactic text were highest in student interest, and audiotapes, videotapes, and LAS were lowest in student interest. The rank order correlation between student high interest and performance across segments-within-media is .86. The rank order correlation between low student interest and performance is - .78.

These results tend to indicate that, on the average, content presented in CAI, audiotape-IP, and syndactic text in this course is both more comprehendable and more interesting to students. Before generalizing these results to all segments within these media, it is necessary to remember that there are, of course, variations within all of the media; i.e., each group of segments-within-media has both reasonably effective and ineffective segments.

The unweighted mean percentage of performance, time, and student interest for both research and non-research segments within each media is given in Appendix U. It can be noted that non-research segments of videotape are the only non-research segments which vary appreciably from research segments-within-media. For videotapes, the non-research segments were somewhat higher (8.7% higher). Implications of these results are that students were not penalized by experimental presentations of material since the research segments on the average were presented just as effectively and interestingly as were non-research segments-within-media.

A second method of presenting data on segments-within-media effectiveness is the percent - percent criterion. The percent - percent criterion is essentially a method of reporting the variability of the distribution of student scores by selecting a criterion point in the distribution of scores and reporting the percentage of students who achieved scores equal to or greater than the criterion (Appendix V).

For purposes of the first implementation, two criterion score points were selected: 70% of the objectives prior to remediation and 80% of the objectives prior to remediation. Additionally, results are reported for the criterion: 80% of the objectives after remediation. Reported in Appendix V are the percent - percent data for all segments-

within-media. Table 7.2 presents a summary of the data in Appendix V for media.

TABLE 7.2

UNWEIGHTED MEAN % STUDENTS SCORING 70% OR 80% ON PROGRESS CHECKS BEFORE AND AFTER REMEDIATION: SEGMENTS-WITHIN-MEDIA

Media	% Students 70% or better before remed.	% Students 80% or better before remed.	% Students 80% or better after remed.
CAI	94.9	81.2	89.8
Audiotape-IP	89.4	74.2	87.5
Syndactic text	80.7	61.7	84.1
Linear text	74.1	53.7	68.0
LAS	61.7	49.2	85.1
Audio-videotape	69.9	44.3	63.9

It can be noted from Table 7.2 that only segments-within-CAI met the average 80%-80% criterion before remediation. However, audiotape-IP and syndactic text segments met the average 80%-80% criterion following remediation. In reviewing these results, it must be remembered that individual segments-within-media and the number of test items per segment influence the

percent-percent averages. For example, three ineffective segments lowered the average percent of students achieving 80% in segments-within-linear text. (See Appendix V: Linear Text, Part Five.) Likewise, too, few items on progress checks influenced the percentage of students responding to 80% of the items. (See Appendix V: Syndactic Text, Segments 5.1, 5.2, and 11.3.) Since fewer than ten items were used in these segments, 80% of the students could miss only one item on the test without automatically lowering the percent-percent criterion to less than 80%-80%.

Initial identification of segments requiring revision was based on the percent-percent distributions. The actual use of percent-percent data in making revisions is explained in Section VIII.

VIII. USE OF EFFECTIVENESS DATA IN COURSE IMPROVEMENT

In addition to being simple indices of course effectiveness, the first implementation performance data were used as the basis for identification and revision of core materials and tests. As discussed in previous sections, the overall results of the first validation were quite good. Since 97% of the students met 70% of the total course objectives without remediation, it is safe to conclude that the materials used were at least minimally effective. On the other hand, since an optimally effective instructional system is one in which all students achieve all objectives, revision of all ineffective segments and test items was undertaken in order to approximate this criterion.

MATERIALS REVISION

Basic steps in the use of performance data for course revision were:

1. The identification of segments within media which fell below 80%, i.e., averaged over media.
2. The identification of each core segment which fell below the 80% - 80% criterion level, i.e., 80% of the students achieving 80% of the objectives.

3. The identification of each objective within segments which was not passed by 80% or more of the students.
4. The identification of student responses to each test item alternative.
5. The identification of student responses to each criterion frame within a programed sequence.

Identification of Media

Results of segments-within-media effectiveness are summarized in Section VII. Results indicated that no media was totally ineffective, but segments of LAS, videotape and audiotape were somewhat less effective and less interesting than CAI, audiotape-IP, and syndactic text.

Since differences in effectiveness between media indicated differences in materials preparation via media rather than differences inherent in the media themselves, the question became "Is it possible to revise materials in these media in such a way that all students will attain all objectives?". The determination was that little could be done to revise LAS segments since they were primarily taught by textbooks, and revision of videotapes (as they were used) was simply not practical.

As a result, the four segments of learning activity summaries were structured into segments of syndactic text, and videotapes were dropped from core segments. (Videotapes of actual leadership situations will be used in depth core segments in the next implementation, but videotapes of simple lectures will not be used in core segments.)

Although audiotapes were somewhat less effective than CAI, audiotape-IP, and syndactic text, they were revised and will be used in the next implementation. Reasons for dropping videotapes and not dropping audiotapes were that a transient presentation was needed for testing hypothesis 1, and audiotapes were less expensive to revise. Audiotapes will be parralleled by a persistent (text) form of presentation in the second implementation, and differences between audiotapes and text presentations will be experimentally studied.

Identification of Segments

Following identification of media which should be eliminated from the system, was identification of ineffective segments, irrespective of media. For example, in three linear text segments, the average percentages of students achieving 80% of the objectives

were 25.0%, 4.5%, and 13.6%. A review of these segments revealed that they should be re-formatted into smaller steps and that lengthy examples should be eliminated. The content outlines for these segments were also revised according to the USNA instructor's recommendations. Other segments requiring substantial revision were identified and reviewed in a similar manner.

Identification of Objectives

In all, 75% of the segments required some revision according to the 80% - 80% criterion. The extent of the revision was based on the percentage of students responding to each objective. For example, a segment falling below the 80% - 80% criterion may be generally ineffective or there may be only two or three objectives which are poorly presented or tested. An example of general ineffectiveness is Part Five, Segment IV, presented by linear text (Table 8.1).

TABLE 8.1

TEST ITEM-OBJECTIVE REFERENCE

(PART FIVE, SEGMENT IV)

Objective identifier	Test item number	Percentage correct responses before remediation	Percentage correct responses after remediation
TO-1	1	65.9	72.7
TO-2	2	84.1	93.2
TO-3	3	29.5	45.5
TO-4	4	79.5	84.1
TO-5	5	100.0	100.0
TO-6	6	90.9	97.7
TO-7	7	25.0	56.8
TO-8	8	47.7	63.6
TO-9	9	79.5	90.9
TO-10	10	93.2	95.5
TO-11	11	61.4	68.2

It can be noted that, before remediation, seven out of 11 text items were answered by less than 80% of the students. Following remediation, five of these items were still below the 80% criterion. The distribution indicated that student performance was generally poor (63% of the objectives were below criterion), and the segment should be revised considerably.

The example of general ineffectiveness can be contrasted with an example of specific ineffectiveness in which 80% of the students perform well on most objectives but do miss two or three of them.

An example of specific ineffectiveness is Part Six, Segment IV, presented by syndactic text (Table 8.2).

TABLE 8.2

TEST ITEM-OBJECTIVE REFERENCE

(PART SIX, SEGMENT IV)

Objective identifier	Test item number	Percentage correct responses before remediation	Percentage correct responses after remediation
TO-1	1	95.5	97.7
TO-2	2	100.0	100.0
TO-3	3	93.2	97.7
TO-4	4	95.5	97.7
TO-5	5	100.0	100.0
TO-7	7	97.7	100.0
TO-9	9	65.9	79.5
TO-10	10	52.3	63.6

Six out of eight items were answered correctly by 90% or more of the students; only two were answered by less than 80% of the students. Therefore, the specific areas of difficulty in this syndactic text segment were the last two objectives, or last two test items. Since students did not improve appreciably on the test items after remediation, and since they remediated through the same core materials, the isolation of specific problems, i.e., test item or materials, is not apparent from the test item-objective reference alone.

Identification of Test Item

The general criterion for revision of either test items or materials was the 80% difficulty level of an item. When items fell below the 80% level, the test item, objective, and content were reviewed with respect to the accuracy of their interrelationship. Where discrepancies occurred, subjective judgment determined if the test item or the materials should be changed in order to achieve a closer correspondence to the objective. Where no discrepancy appeared, the following procedures were implemented:

The distribution of responses to each distractor was studied in relation to the materials in order to determine a) if the distractors were in fact correct, b) if the

distractors were incorrect but not presented as distinctly incorrect, or c) if the correct answer was not emphasized as correct.

Revision based on these considerations generally consisted of strengthening the controlling stimulus for the correct response by increasing the similarity of examples in the materials and the example used in the text item, or vice versa. Where incorrect distractors were partially or actually correct, subject matter experts made decisions on revising the content outline or material.

When available, the discrimination index of each item was used in conjunction with the difficulty level in deciding revision. Although the difficulty level was weighed more heavily in revision, all negatively discriminating items, regardless of difficulty indicated need for revision. Since negative discrimination indicated that students who scored high or all other items were failing an item, the materials were reviewed to see if in fact the correct answer was the only correct response or if there was another distractor which was also correct. If the difficulty level was 90% or above, a discrimination index was not considered, since only two or three students could cause the discrimination index to be negative. However, if the difficulty ranged

between 80 and 90 and the discrimination was negative, the need for revision to either test item or materials was clear.

Although the discrimination indices were not available until after many test items and materials were revised, all progress checks were checked to determine if items that were negatively discriminating or significantly positively discriminating were revised under the difficulty rule. General guidelines for use of both difficulty and discrimination in revision are given in Table 8.3.

TABLE 8.3

REVISION DECISIONS BASED ON PROGRESS CHECK DATA

Range of difficulty	Discrimination	Revision	Test item	Material
90-100	+	No	No	No
	—	No	No	No
	o	No	No	No
80-90	+	No	No	No
	—	Yes	?	?
	o	No	No	No
60-80	+	Yes	?	Yes
	—	Yes	Yes	Yes
	o	Yes	?	?
0-60	+	Yes	?	?
	—	Yes	?	?
	o	Yes	?	?

Since data on item discrimination was not available until late in the revision cycle, it was not possible to fully implement these decision rules and verify their usefulness in the revision process. However, in subsequent revisions, these data will be used as additional guidelines to writers.

Identification of Frames

When materials required revision, writers analyzed material by criterion frames within the program sequences. When test items required revision, writers revised items according to the item response analysis, item difficulty and item discrimination index. In materials revision, frame analysis indicated which criterion objectives included in the program sequence were ineffective. (An example of the frame analysis format for syndactic text is given in Appendix W.) High and low error rates for both summary quizzes and frames of syndactic text were indicated on the frame analysis sheet. From these data, the writer could determine whether prerequisite or terminal behaviors were failed by most students.

Theoretical decision rules for frame revision were supplemented by subjective judgment concerning the appropriateness of examples and clarity of presentation being taught in weak frames.

TEST ITEM REVISION

Revision of progress checks was based on the 80% difficulty level, the item discrimination index, and the item response analysis. Since these procedures have already been indicated in Table 8.3, this section is limited to revision procedures for the administrative and cumulative posttests.

Administrative Test Revision

The following guidelines were given to writers for the first revision of administrative test items:

1) Identify and revise all items which are above 75% difficulty on the pretest.
2) Identify and revise all items which have less than 75% difficulty if the difficulty of the corresponding posttest item is less than 20% above that of the pretest item.
3) Identify and revise items if the posttest difficulty is less than 70%.
4) Under condition 2 above, revise all items which have negative discrimination indices or which have positive discrimination indices that are significant.
5) Wherever possible, revise administrative items at the same time the corresponding materials and progress check items are revised, and revise materials if they are the cause of negative discrimination or low difficulty.

After all administrative items had been revised and were reviewed by WLC subject matter experts, the items were pretested on two groups of USNA midshipmen, plebes and second classmen. Plebes were tested in order to determine which items were simply too easy. Second classmen were tested in order to determine which items measured objectives that second classmen had already learned.

Specifications given to writers for the second revision of administrative test items were:

1) Work from second class to plebe data. Identify all items which 75% or more second classmen answered correctly. If 50% or more plebes also answered correctly, it is simply too easy and needs to be revised or replaced. If less than 50% of the plebes answered correctly, leave it alone unless 80% or more of the second classmen answered correctly. In the latter case, replace the item.
2) Work from plebe to second class data. Identify all items which 50% or more plebes answered correctly. If 66% or more second classmen also answered correctly, it is too easy and needs revision.

3) Use confidence data where it appears appropriate, i.e., if in doubt as to necessity for revision, check the confidence data for second classmen. The percentage of students getting the item correct but indicated no confidence in their response is the percentage who were guessing. However, confidence data should in no way suggest necessity for revision which the item response data does not suggest.

Cumulative Posttest Revision

Since CPT items were constructed to discriminate among students and among modules and since there was no pretest for CPTs, the guidelines for revision of CPTs differed from those for the administrative tests. Guidelines for CPT revision were:

1) Identify all items which are over 85% difficulty or below 40% difficulty. These items will usually need to be revised considerably or replaced.
2) Identify items which are between 75% - 85% difficulty if they contain one distractor which has elicited no responses, and revise that distractor.
3) Identify items which are between 65% - 75% difficulty if they contain two distractors

which elicit no responses, and revise either of the distractors.

4) Identify items between 40% - 50% if they contain a distractor which elicits more correct responses than the actual correct response, and revise either the distractor or the correct alternatives.

5) Identify and revise items between the 40% - 80% difficulty level that are negatively discriminating and were not revised according to the difficulty rules.

6) If an item has a significant positive discrimination index, do not revise it.

REFERENCES

Aerospace Education Foundation. Technology and Innovation in Education. New York: Frederick A. Praeger, Inc., 1968.

Anderson, R. C., et al. (Ed.). Current Research on Instruction. Englewood Cliffs: Prentice-Hall, 1969.

Bloom, B. S. (Ed.). Taxonomy of Educational Objectives. New York: Longmans, Green, 1954.

Brennan, R. L. Some statistical problems in the evaluation of self-instructional programs. Unpublished doctoral dissertation, Harvard University, 1970.

Deterline, W. A. Practical problems in program production. The Sixty-sixth Yearbook of the N.S.S.E., 1967, Part II, 178-216.

Ellis, H. C. Research design in programmed learning. National Society for Programmed Instruction. August 1962, 7-8.

Evans, J. L., Glaser, R., and Homme, L. E. "The RULEG System for the Construction of Programmed Verbal Learning Sequences," 1960.

Gagne, R. M. The Conditions of Learning. New York: Holt, Rinehart and Winston, 1965.

Horst, P. Psychological Measurement and Prediction. Belmont, California: Wadsworth Publishing Company, Inc., 1966.

Leadership Management Course. Sequencing Rationale. TP-6.2, Contract No. N00600-68-C-1525, March 21, 1969.

Leadership Management Course. Research and Evaluation Plans - Part I. TR-6.3a, Contract No. N00600-68-C-1525, January 31, 1969.

Leadership Management Course. Research and Evaluation Plans - Part II. TR-6.3a, Contract No. N00600-68-C-1525, April 30, 1969.

Leadership Management Course. Instructor's Guide. TR-6.6, Contract No. N00600-68-C-1525,

Leadership Management Course. Student Guide. TR-6.5, Contract No. N00600-68-C-1525,

Mager, R. F. Preparing Instructional Objectives. Palo Alto: Fearon Publishers, Inc., 1962.

Markle, S. M. Good Frames and Bad. New York: John Wiley and Sons, 1964.

McNemar, Q. On growth measurement. Educational and Psychological Measurement, 1958, 18, 47-55.

Report of Phase II Research Findings: The Design and Methodology for Research on the Interaction of Media, Conditions of Instruction, and Student Characteristics for a Multi/Media Course in Leadership, Psychology and Management, Part I: Conditions of Instruction. TR-6.12a. Contract No. N00600-68-C-1525, October, 1970.

Report of Phase II Research Findings: The Design and Methodology for Research on the Interaction of Media, Conditions of Instruction, and Student Characteristics for a Multi/Media Course in Leadership, Psychology and Management, Part II: Relationship of Student Characteristics to Media and Conditions of Instruction. TR-6.12b. Contract No. N00600-68-C-1525, October, 1970.

Rosenthal, R. Experimenter Effects in Behavioral Research. New York: Appleton-Century-Crofts, 1966.

Schramm, W. The Research on Programed Instruction. Washington, D. C.: U. S. Department of Health, Education, and Welfare, 1964. (OE-34034).

Stolurow, L. M. Instructional systems validation planning. Technical recommendation No. 2. USNA Contract No. N00161-7339-4781, March 29, 1968.

Tosti, D. T. and Ball, J. R. A Behavioral Approach to Instructional Design and Media Selection. Albuquerque, 1968: Behavior Systems Division, Westinghouse Learning Corporation.

APPENDIX A

COURSE OUTLINE

APPENDIX A

COURSE OUTLINE

PART ONE: Overview of Leadership

Segment I	Concepts of Leadership
Segment II	Standards of Leadership in the Naval Service

PART TWO: Individual Behavior

Segment I	Introduction to Psychology
Segment II	Behavior and Its Observation
Segment III	Learning
Segment IV	Factors Affecting Learning
Segment V	Attention and Perception
Segment VI	Motivation
Segment VII	Conflict
Segment VIII	Neurotic and Psychotic Reactions
Segment IX	Personality

PART THREE: Group Dynamics

Segment I	Characteristics of Groups
Segment II	The Relation of the Leader to the Group
Segment III	Group Interactions
Segment IV	Conformity as a Factor of Group Behavior
Segment V	Relation of the Individual to the Group

PART FOUR: Achieving Effective Communication

Segment I	Importance of Interpersonal Communication
Segment II	Types of Communication
Segment III	The Communication Process (Receiver and Barriers)
Segment IV	The Communication Process (Sender and Feedback)
Segment V	Formal Communication and Its Dimensions
Segment VI	Informal Communication
Segment VII	Communication Under Battle Situations

PART FIVE: Military Management

Segment I	Introduction to Management and the Management Process
Segment II	Decision Making and Creativity
Segment III	Objectives
Segment IV	Planning
Segment V	Organizing: Principles and Process
Segment VI	Organizing: Structure
Segment VII	Organizing: Charting
Segment VIII	Directing
Segment IX	Controlling
Segment X	Coordinating

PART SIX: Authority and Responsibility

Segment I	Concept of Authority
Segment II	Why People Accept/Resist Authority
Segment III	Delegation of Authority; Line-Staff Relationship
Segment IV	Responsibility

PART SEVEN: Leadership Behavior and Style

Segment I	Leadership Behavior
Segment II	Leadership Style
Segment III	Determiners of Leadership Style - The Leader
Segment IV	Determiners of Leadership Style - The Group and The Situation
Segment V	Participative Leadership

PART EIGHT: Senior-Subordinate Relationships

Segment I	Organizational Structure & Social Distance in Senior-Subordinate Relationships
Segment II	Officer-Enlisted Relationships
Segment III	Assumption of Command and Formal & Informal Leader Relationships
Segment IV	Introduction to Counseling
Segment V	The Counseling Process
Segment VI	Relations with Seniors and Contemporaries

PART NINE: Morale -- Esprit de Corps

Segment I	Morale
Segment II	Group Solidarity and Esprit

PART TEN: Discipline

Segment I	Introduction to Discipline
Segment II	Development and Maintenance of Discipline

PART ELEVEN: Personnel Evaluation

Segment I	The Role of Evaluation
Segment II	Enlisted Performance Evaluation
Segment III	Officer Evaluation

PART TWELVE: Applied Leadership

Segment I	Measurement of Effective Leadership
Segment II	Generally Recognized Characteristics of an Effective Leader
Segment III	Techniques of Assuming Command
Segment IV	"That's an Order!"

APPENDIX B

GUIDELINES FOR THE DEVELOPMENT OF BEHAVIORAL OBJECTIVES

APPENDIX B

GUIDELINES FOR DEVELOPMENT OF BEHAVIORAL OBJECTIVES

A. Type 1 -- Definition - Identification

Given the instruction to identify the correct purpose of (definition of/description of/use of) concept X, the M will select from several choices the correct purpose of (definition of/description of/use of) concept X.

1. General Type 1
   a. Given the instruction to define concept X,
   b. The M will select from several choices a definition of X
   c. Similar to the following: "X...."

   (NOTE: The third part is optional.)
2. Example of Type 1
   a. When given the instruction to define "acquisition,"
   b. The M will select from several choices the correct definition.
3. Example of Type 1
   a. Given several choices, the M will select the correct definition of acquisition.

4. Example of Type 1

   a. When asked to define "attention,"

   b. The M will select from several choices a statement

   c. (which indicates that attention is the selection of specific stimulus elements).

B. Type 2 -- Discrimination - Comparison

Given the instruction to evaluate the relationship between (defining attr butes of/contrast between/comparison among) classes X, Y, Z...N, the M will be able to select from several choices the paragraph which illustrates (describes/differentiates/identifies) the relationship between (defining attributes of/contrast between/comparison among) classes X, Y, Z...N.

1. General Type 2

   a. When asked to evaluate the relationship among X, Y,...N,

   b. The M will select from several choices the paragraph which describes this relationship.

2. Example of Type 2

   a. When asked to evaluate the difference between retroactive and proactive inhibition,

   b. The M will select from several choices the paragraph which describes this difference.

C. Type 3 -- Generalization - Problem Identification

1. Deductive

   a. Given examples of X, the M will be able to select from several choices the example which illustrates principle Y.

2. Inductive

   a. Given an example of X, the M will be able to select from several choices which principle (X, Y,...or N) is shown (exemplified/demonstrated) by the example.

3. General Type 3

   a. Given examples of an X,

b. The M will be able to select from several choices the example which illustrates principle Y.

4. Example of Type 3

a. When asked to compare several versions of the same communication,

b. The M will select from several choices the version which clearly links the subordinate's role to the overall objective of Naval operations.

D. Type 4 -- Problem Solving

When asked to evaluate a situation which is an example of class X, the M will select from several choices the correct solution (approach/method/resolution of/reaction to) the situation, using principle Y.

1. General Type 4

a. Given a problem situation which is an example of class X,

b. The M will select from several choices the correct approach to the situation,

c. Using principle Y.

(NOTE: The third part may be omitted if the objective is unambiguous. In practive, the third part is usually not given to the student.)

2. Example of Type 4

a. When asked to evaluate a situation in which there is an apparent failure in communication,

b. The M will select from several choices the description which indicates the appropriate action of a leader

c. Who assumes responsibility for the failure.

APPENDIX C

OUTLINE OF COURSE STRUCTURE AND MEDIA

Appendix C

OUTLINE OF COURSE STRUCTURE AND MEDIA

Part and Segment Number	Content Heading	CPT Unit[a]	Medium[b]
	PART ONE: OVERVIEW OF LEADERSHIP		
1.1	Concepts of Leadership	NR	ST
1.2	Standards of Leadership in the Naval Service	NR	F-GD
	PART TWO: INDIVIDUAL BEHAVIOR		
2.1	Introduction to Psychology	NR	ST
2.2	Behavior and its Observation	1	AT- or VT-PB
2.3	Learning	1	AT- or VT-PB
2.4	Factors Affecting Learning	1	AT- or VT-PB
2.5	Attention and Perception	1	AT- or VT-PB
2.6	Motivation	2	ST
2.7	Conflict	2	ST
2.8	Neurotic and Psychotic Reactions	2	ST
2.9	Personality	NR	LAS
	PART THREE: GROUP DYNAMICS		
3.1	Characteristics of Groups	3	AT- or VT-PB
3.2	The Relationship of the Leader to the Group	3	AT- or VT-PB
3.3	Group Interactions	3	AT- or VT-PB
3.4	Conformity as a Factor of Group Behavior	3	AT- or VT-PB
3.5	Relation of the Individual to the Group	NR	ST
	PART FOUR: ACHIEVING EFFECTIVE COMMUNICATION		
4.1	Importance of Interpersonal Communication	4	LT
4.2	Types of Communication	4	LT
4.3	The Communication Process (Receiver and Barriers)	4	LT
4.4	The Communication Process (Sender and Feedback)	5	AT- or VT-PB
4.5	Formal Communication and Its Dimensions	5	AT- or VT-PB
4.6	Informal Communication	5	AT- or VT-PB
4.7	Communication Under Battle Situations	5	AT- or VT-PB
	PART FIVE: MILITARY MANAGEMENT		
5.1	Introduction to Management and the Management Process	NR	ST
5.2	Decision Making and Creativity	NR	ST
5.3	Objectives	NR	ST
5.4	Planning	6	LT
5.5	Organizing: Principles and Process	6	LT
5.6	Organizing: Structure	6	LT
5.7	Organizing: Charting	7	AT- or VT-PB

Part and Segment Number	Content Heading	CPT Unit[a]	Medium
	PART FIVE: MILITARY MANAGEMENT (CON'T)		
5.8	Directing	7	AT- or VT-PB
5.9	Controlling	7	AT- or VT-PB
5.10	Coordinating	7	AT- or VT-PB
	PART SIX: AUTHORITY AND RESPONSIBILITY		
6.1	Concept of Authority	8	ST
6.2	Why People Accept/Resist Authority	8	ST
6.3	Delegation of Authority; Line-Staff Relationship	8	ST
6.4	Responsibility	NR	ST
	PART SEVEN: LEADERSHIP BEHAVIOR AND STYLE		
7.1	Leadership Behavior	9	AT- or VT-PB
7.2	Leadership Style	9	AT- or VT-PB
7.3	Determiners of Leadership Style - The Leader	9	AT- or VT-PB
7.4	Determiners of Leadership Style - The Group and The Situation	9	AT- or VT-PB
7.5	Participative Leadership	NR	VT-PB
	PART EIGHT: SENIOR-SUBORDINATE RELATIONSHIPS		
8.1	Organizational Structure & Social Distance in Senior-Subordinate Relationships	10	LT
8.2	Officer-Enlisted Relationships	10	LT
8.3	Assumption of Command and Formal & Informal Leader Relationships	10	LT
8.4	Introduction to Counseling	11	LAS
8.5	The Counseling Process	11	LAS
8.6	Relations with Seniors and Contemporaries	11	LAS
	PART NINE: MORALE - ESPRIT DE CORPS		
9.1	Morale	NR	VT-PB
9.2	Group Solidarity and Esprit	NR	VT-PB
	PART TEN: DISCIPLINE		
10.1	Introduction to Discipline	NR	AT-IP
10.2	Development and Maintenance of Discipline	NR	AT-IP
	PART ELEVEN: PERSONNEL EVALUATION		
11.1	The Role of Evaluation	12	ST
11.2	Enlisted Performance Evaluation	12	ST
11.3	Officer Evaluation	12	ST

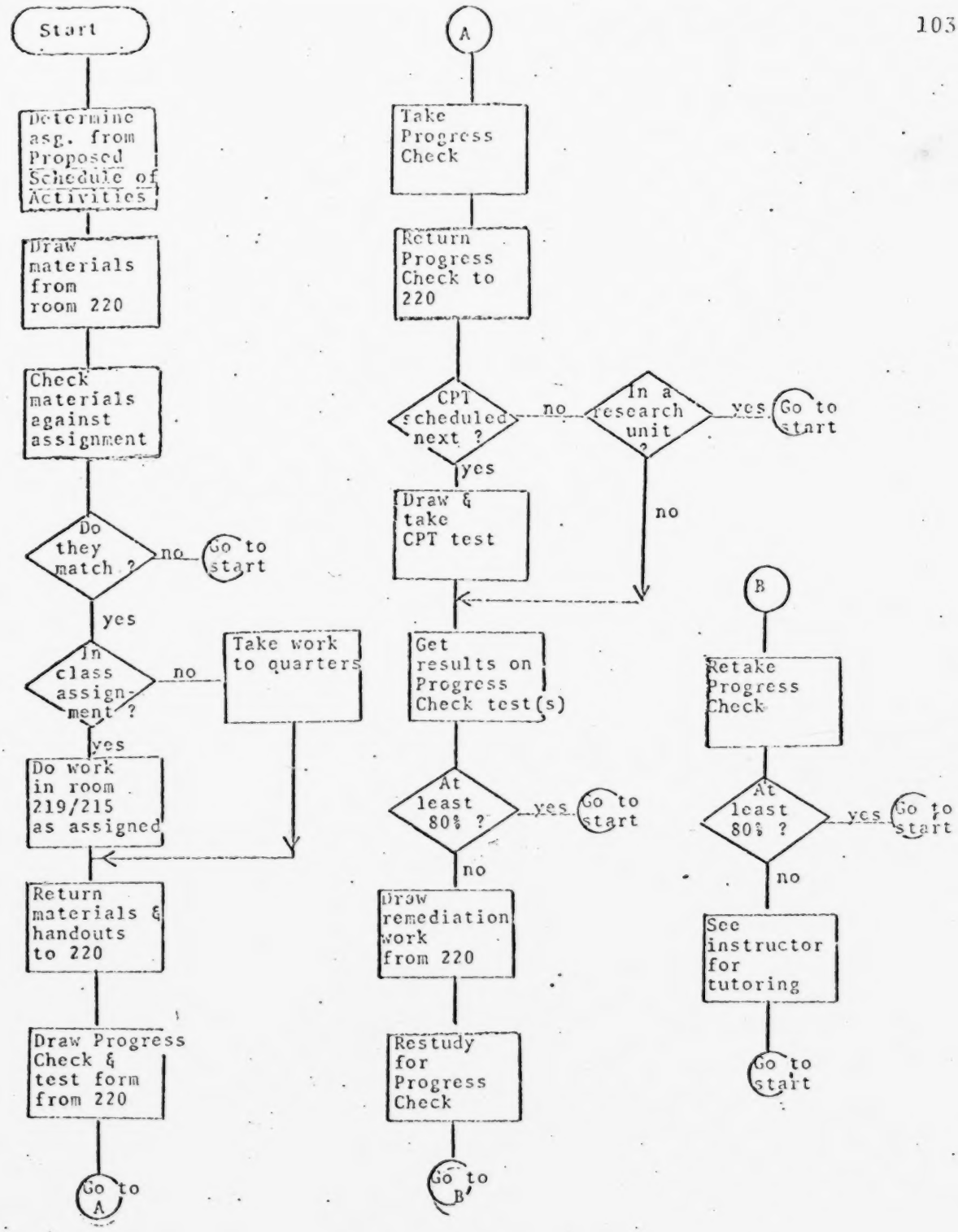
Part and Segment Number	Content Heading	CPT Unit[a]	Medium
	PART TWELVE: APPLIED LEADERSHIP		
12.1	Measurement of Effective Leadership	13	CAI
12.2	Generally Recognized Characteristics of an Effective Leader	13	CAI
12.3	Techniques of Assuming Command	13	CAI
12.4	"That's an Order!"	13	CAI

a NR refers to a nonresearch segment, thus not assigned to a CPT unit.

b ST=Syndactic (multi-level) Text; F-GD=Film, Group Discussion; AT=Audiotape; VT=Videotape; PB=Panelbook; LAS=Learning Activities Summary; LT=Linear Text; IP=Intrinsic Program; CAI=Computer Assisted Instruction.

APPENDIX D

FLOWCHART OF THE STUDENT'S INSTRUCTIONAL SEQUENCE



APPENDIX-D

FLOW CHART OF THE INSTRUCTIONAL SEQUENCE

APPENDIX E

EXAMPLE: BEHAVIORAL OBJECTIVE AND CRITERION REFERENCED TEST ITEM

APPENDIX E

EXAMPLE OF BEHAVIORAL OBJECTIVE:

Given examples of leaders interacting with groups and using various leadership styles, the M will be able to identify the situation in which the leader uses participative leadership.

EXAMPLE OF CRITERION REFERENCED TEST ITEM:

LT Henry was the type of officer who left no doubt in the minds of his subordinates as to his leadership qualifications. He was always in the area, insuring that his orders were being carried out. When there was time, he consulted with his leading petty officers to obtain their ideas and recommendations and shared decision making with them. LT Henry made sure that credit was given to those who deserved it. LT Henry took the responsibility for success or failure.

From the choices below, identify the type of leadership style being used by LT Henry.

a. Participative
b. Supervisory
c. Authoritarian
d. Shared-leadership

APPENDIX F

FREQUENCY DISTRIBUTION OF ITEM DIFFICULTY AND DISCRIMINATION: ADMINISTRATIVE POSTTEST

APPENDIX F

ITEM DISCRIMINATION AND ITEM DIFFICULTY FREQUENCY DISTRIBUTION SUMMARY (POSTTEST)

INTERVAL RANGE	ITEM DISCRIMINATION	INTERVAL RANGE	ITEM DISCRIMINATION	ITEM DIFFICULTY
-0.200 TO -0.250	0	0.0 TO 0.049	1	1
-0.249 TO -0.200	0	0.050 TO 0.099	7	0
-0.199 TO -0.150	0	0.100 TO 0.149	1	3
-0.149 TO -0.100	0	0.150 TO 0.199	7	0
-0.099 TO -0.050	1	0.200 TO 0.249	6	0
-0.049 TO -0.000	0	0.250 TO 0.299	3	0
-0.699 TO -0.650	0	0.300 TO 0.349	3	1
-0.649 TO -0.600	0	0.350 TO 0.399	2	3
-0.599 TO -0.550	2	0.400 TO 0.449	1	1
-0.549 TO -0.500	0	0.450 TO 0.499	2	2
-0.499 TO -0.450	0	0.500 TO 0.549	3	6
-0.449 TO -0.400	0	0.550 TO 0.599	1	1
-0.399 TO -0.350	0	0.600 TO 0.649	1	3
-0.349 TO -0.300	2	0.650 TO 0.699	0	5
-0.299 TO -0.250	4	0.700 TO 0.749	1	4
-0.249 TO -0.200	4	0.750 TO 0.799	0	9
-0.199 TO -0.150	3	0.800 TO 0.849	7	2
-0.149 TO -0.100	3	0.850 TO 0.899	1	7
-0.099 TO -0.050	2	0.900 TO 0.949	0	9
-0.049 TO 0.000	5	0.950 TO 0.999	0	12
0.000 - 0.000	6	1.000 to ----		6

APPENDIX G

EXAMPLE: ITEM DISCRIMINATION STATISTICS

APPENDIX G

ITEM DISCRIMINATION AND RELIABILITY ANALYSIS

ADMINISTRATIVE POSTTEST

ITEM 1

TOTAL SCORE	FREQUENCIES CORRECT	FREQUENCIES INCORRECT	FREQUENCIES TOTAL	CUMULATIVE FREQUENCIES CORRECT	CUMULATIVE FREQUENCIES INCORRECT	CUMULATIVE FREQUENCIES TOTAL
75	0	0	0	29	15	44
74	0	0	0	29	15	44
73	0	0	0	29	15	44
72	0	0	0	29	15	44
71	0	0	0	29	15	44
70	0	0	0	29	15	44
69	0	0	0	29	15	44
68	0	0	0	29	15	44
67	0	0	0	29	15	44
66	0	0	0	29	15	44
65	0	0	0	29	15	44
64	1	0	1	29	15	44
63	1	0	1	28	15	43
62	0	0	0	27	15	42
61	1	0	1	27	15	42
60	1	2	3	26	15	41
59	3	3	6	25	13	38
58	3	0	3	22	10	32
57	0	2	2	17	10	27
56	1	2	3	17	8	25
55	6	2	8	16	6	22
54	2	2	4	10	4	14
53	2	0	2	8	2	10
52	1	1	2	6	2	8
51	1	0	1	5	1	6
50	0	0	0	4	1	5
49	0	0	0	4	1	5
48	2	1	3	4	1	5
47	1	0	1	2	0	2
46	1	0	1	1	0	1
45	0	0	0	0	0	0
44	0	0	0	0	0	0
43	0	0	0	0	0	0
42	0	0	0	0	0	0
41	0	0	0	0	0	0
40	0	0	0	0	0	0
39	0	0	0	0	0	0
38	0	0	0	0	0	0
37	0	0	0	0	0	0
36	0	0	0	0	0	0
35	0	0	0	0	0	0
34	0	0	0	0	0	0
33	0	0	0	0	0	0
32	0	0	0	0	0	0
31	0	0	0	0	0	0
30	0	0	0	0	0	0
29	0	0	0	0	0	0
28	0	0	0	0	0	0
27	0	0	0	0	0	0
26	0	0	0	0	0	0
25	0	0	0	0	0	0
24	0	0	0	0	0	0
23	0	0	0	0	0	0
22	0	0	0	0	0	0
21	0	0	0	0	0	0
20	0	0	0	0	0	0
19	0	0	0	0	0	0
18	0	0	0	0	0	0
17	0	0	0	0	0	0
16	0	0	0	0	0	0
15	0	0	0	0	0	0
14	0	0	0	0	0	0
13	0	0	0	0	0	0
12	0	0	0	0	0	0
11	0	0	0	0	0	0
10	0	0	0	0	0	0
9	0	0	0	0	0	0
8	0	0	0	0	0	0
7	0	0	0	0	0	0
6	0	0	0	0	0	0
5	0	0	0	0	0	0
4	0	0	0	0	0	0
3	0	0	0	0	0	0
2	0	0	0	0	0	0
1	0	0	0	0	0	0
0	0	0	0	0	0	0

DIFFICULTY P	Q	P-Q	P(MAX)	INDEX R	VARIANCE OF P-Q	STANDARD ERROR OF P-Q	CORRECTION FOR CONTINUITY	TEST STATISTIC	N1	NO
0.659	170 225	-47	47.4	-0.11	6449.306	80.309	2.067	0.559	29	15

APPENDIX H

TEST ITEM STATISTICS SUMMARY: ADMINISTRATIVE POSTTEST

APPENDIX H

TEST ITEM STATISTICS SUMMARY (POSTTEST)

TEST ITEM ID	POSTTEST ITEM ID	ITEM DIFFICULTY	ITEM DISCRIMINATION	SIGNIFICANCE	TEST ITEM ID	POSTTEST ITEM ID	ITEM DIFFICULTY	ITEM DISCRIMINATION	SIGNIFICANCE
3003010	3303028	0.614	0.25	1.344	3005922	3305970	0.955	-0.32	0.664
3001210	3301246	0.773	0.07	0.296	3002819	3302818	0.591	0.36	1.927
3005007	3305014	0.932	0.16	0.562	3004522	3304552	0.864	0.07	0.245
3003503	3305519	1.000	** -1.00	0.0	3003726	3303709	0.932	0.57	1.877
3001008	3301055	0.977	0.81	1.280	3004219	3304277	0.864	0.49	1.977
3001312	3301364	0.545	0.10	0.506	3005508	3305520	0.909	-0.05	0.028
3000122	3300133	0.500	0.03	0.141	3003922	3302957	0.750	0.18	0.697
3005401	3305456	0.977	-0.16	0.194	3003220	3303202	0.932	-0.28	1.047
3002110	3302172	0.636	0.25	0.787	3004721	3304704	1.000	** -1.00	0.0
3000607	3300671	0.818	-0.09	0.256	3003019	3303075	0.545	0.52	1.770
3004511	3304566	0.659	0.25	1.295	3000805	3300801	0.932	0.52	1.630
3005713	3305740	0.977	0.21	0.243	3000425	3300613	0.750	-0.20	1.491
3001411	3301436	0.886	-0.13	0.666	3002205	3302244	0.795	-0.12	0.581
3003634	3303672	0.955	0.36	0.657	3000813	3300960	0.682	0.63	2.313
3000912	3300960	0.773	-0.02	0.047	3002916	3302912	0.705	0.56	1.960
3005113	3305129	1.000	** -1.00	0.0	3003012	3305063	0.977	-0.79	1.938
3002905	3302925	0.932	0.54	1.484	3004322	3304324	0.636	0.52	1.636
3000206	3300237	0.818	-0.10	0.406	3000610	3300643	0.932	-0.58	1.573
3004410	3304458	0.864	0.16	0.538	3002024	3302021	0.455	0.06	0.208
3001818	3301876	0.955	-0.67	1.279	3002021	3302006	0.977	-0.05	1.938
3002716	3302775	0.682	-0.20	1.025	3005010	3305036	0.705	0.03	0.139
3004210	3304226	0.841	0.17	0.673	3002614	3302627	0.977	0.60	0.956
3004021	3304045	0.977	0.21	1.280	3001418	3301430	0.932	0.20	0.409
3005512	3305525	0.977	-0.09	-0.054	3004023	3304041	0.750	0.28	1.252
3005912	3305947	0.545	0.07	0.334	3002519	3302505	1.000	** -1.00	0.0
3003112	3303180	0.886	-0.22	1.176	3005617	3305678	0.705	-0.06	0.376
3002812	3302806	0.886	0.06	0.312	3000217	3300225	0.045	-0.15	0.287
3003710	3303749	0.864	-0.32	1.207	3004913	3304907	1.000	** -1.00	0.0
3001706	3301776	0.955	0.74	1.681	3001109	3301139	0.864	0.14	0.493
3000503	3300511	0.932	0.16	0.377	3003313	3303307	0.682	0.28	1.443
3005605	3305669	1.000	** -1.00	0.0	3005409	3305459	0.750	0.09	0.377
3004614	3304615	0.477	-0.02	0.050	3001615	3301634	0.955	-0.22	1.186
3003205	3303264	0.614	0.49 *	2.000	3005517	3305561	0.773	-0.02	0.081
3001509	3301568	0.705	-0.20	0.962	3001824	3301803	0.136	-0.25	0.947
3005207	3305219	0.932	0.02	0.007	3001315	3301331	0.114	-0.28	0.965
3000317	3300310	0.864	0.55	2.099	3000516	3300517	0.114	-0.21	0.830
3004115	3304121	0.795	0.19	0.647	3000104	3300142	0.659	-0.26	1.336
3001313	3301355	0.500	0.32	1.309					

* Significant discrimination

** Zero discrimination

APPENDIX I

EXAMPLE: ITEM RESPONSE ANALYSIS: ADMINISTRATIVE POSTTEST

POST-TEST ITEM RESPONSE ANALYSIS

NUMBER OF ITEMS=75

ITEM NO.	POST ITEM I.D.		ITEM RESPONSE FD	ITEM RESPONSE PC	CONFIDENCE RATINGS 0 FD / PC	1 FD / PC	2 FD / PC	3 FD / PC	4 FD / PC	5 FD / PC	6 FD / PC	7 FD / PC	8 FD / PC	9 FD / PC	CONFIDENCE NCOUNTS
6	3003007														
		* A	30	68.2	0 / 0.0	0 / 0.0	0 / 0.0	0 / 0.0	1 / 3.7	0 / 0.0	4 / 10.0	5 / 10.5	9 / 33.0	8 / 20.0	27
		B	0	0.0	0 / 0.0	0 / 0.0	0 / 0.0	0 / 0.0	0 / 0.0	0 / 0.0	0 / 0.0	0 / 0.0	0 / 0.0	0 / 0.0	0
		C	2	4.5	0 / 0.0	0 / 0.0	0 / 0.0	0 / 0.0	0 / 0.0	0 / 0.0	0 / 0.0	1 / 50.0	0 / 0.0	1 / 50.0	2
		D	12	27.3	0 / 0.0	0 / 0.0	0 / 0.0	1 / 10.0	0 / 0.0	0 / 0.0	1 / 10.0	1 / 10.0	4 / 40.0	3 / 30.0	10
	NCOUNTS		44		0	0	0	1	1	0	5	7	13	12	39
								NON-RESPONDERS= 0							
10	3003008														
		A	1	2.3	0 / 0.0	0 / 0.0	0 / 0.0	0 / 0.0	0 / 0.0	0 / 0.0	0 / 0.0	0 / 0.0	0 / 0.0	1 / 100.0	1
		* B	43	97.7	0 / 0.0	0 / 0.0	0 / 0.0	1 / 2.6	0 / 0.0	1 / 2.6	5 / 13.2	5 / 13.2	9 / 23.7	17 / 44.7	38
		C	0	0.0	0 / 0.0	0 / 0.0	0 / 0.0	0 / 0.0	0 / 0.0	0 / 0.0	0 / 0.0	0 / 0.0	0 / 0.0	0 / 0.0	0
		D	0	0.0	0 / 0.0	0 / 0.0	0 / 0.0	0 / 0.0	0 / 0.0	0 / 0.0	0 / 0.0	0 / 0.0	0 / 0.0	0 / 0.0	0
	NCOUNTS		44		0	0	0	1	0	1	5	5	9	18	39
								NON-RESPONDERS= 0							
37	3003709														
		A	0	0.0	0 / 0.0	0 / 0.0	0 / 0.0	0 / 0.0	0 / 0.0	0 / 0.0	0 / 0.0	0 / 0.0	0 / 0.0	0 / 0.0	0
		* B	41	93.2	0 / 0.0	0 / 0.0	0 / 0.0	2 / 5.6	0 / 0.0	0 / 0.0	3 / 8.3	4 / 11.1	10 / 27.8	17 / 47.2	36
		C	3	6.8	0 / 0.0	0 / 0.0	0 / 0.0	0 / 0.0	0 / 0.0	0 / 0.0	1 / 33.3	1 / 33.3	0 / 0.0	1 / 33.3	3
		D	0	0.0	0 / 0.0	0 / 0.0	0 / 0.0	0 / 0.0	0 / 0.0	0 / 0.0	0 / 0.0	0 / 0.0	0 / 0.0	0 / 0.0	0
	NCOUNTS		44		0	0	0	2	0	0	4	5	10	18	39
								NON-RESPONDERS= 0							

APPENDIX J

FREQUENCY DISTRIBUTION OF ITEM DIFFICULTY AND POSITIVE DISCRIMINATION INDICES: PROGRESS CHECKS

APPENDIX J

FREQUENCY DISTRIBUTION OF ITEM DIFFICULTY
AND POSITIVE DISCRIMINATION INDICES: PROGRESS CHECKS

Interval range	Positive discrimination frequency	Difficulty frequency
1.000-		69
.950-.999	1	80
.900-.949	3	67
.850-.899	1	48
.800-.849	2	37
.750-.799	4	56
.700-.749	3	30
.650-.699	4	27
.600-.649	7	9
.550-.599	8	12
.500-.549	4	17
.450-.499	11	16
.400-.449	13	8
.350-.399	24	10
.300-.349	19	8
.250-.299	30	12
.200-.249	28	5
.150-.199	37	6
.100-.149	30	5
.050-.099	27	5
.000-.049	31	3

APPENDIX K

FREQUENCY DISTRIBUTION OF NEGATIVE ITEM DISCRIMINATION INDICES

APPENDIX K

FREQUENCY DISTRIBUTION OF NEGATIVE ITEM DISCRIMINATION INDICES

Interval range	Negative discrimination frequency
-.999 to -.950	1
-.949 to -.900	
-.899 to -.850	1
-.849 to -.800	3
-.799 to -.750	1
-.749 to -.700	1
-.699 to -.650	1
-.649 to -.600	1
-.599 to -.550	1
-.549 to -.500	4
-.499 to -.450	4
-.449 to -.400	6
-.399 to -.350	10
-.349 to -.300	12
-.299 to -.250	12
-.249 to -.200	13
-.199 to -.150	17
-.149 to -.100	14
-.099 to -.050	35
-.049 to -.000	30

APPENDIX L

FREQUENCY DISTRIBUTION OF ADMINISTRATIVE PRE AND POSTTEST SCORES

APPENDIX L

FREQUENCY DISTRIBUTION OF ADMINISTRATIVE PRE AND POSTTEST SCORES

Scores	Pretest frequency	Posttest frequency
60-64		6
55-59		24
50-54	1	9
45-49	7	5
40-44	18	
35-39	11	
30-34	6	
25-29	1	

APPENDIX M

GAIN RATIOS FOR ALL STUDENTS BASED ON ADMINISTRATIVE PRE AND POSTTESTS

APPENDIX M

GAIN RATIOS FOR ALL STUDENTS BASED ON ADMINISTRATIVE PRE AND POSTTESTS

Students	Pretest score	Posttest score	Crude gain	Maximum possible gain	Crude gain ratio	Estmtd. true pretest	Estmtd. true gain	Estmtd. maximum true gain	Estmtd. true gain ratio
1	45	64	19	30	63.3	42.7	14.1	32.3	43.6
2	44	56	12	31	38.7	42.2	14.8	32.8	45.2
3	31	52	21	44	47.7	35.4	17.1	39.6	43.3
4	35	58	23	40	57.5	37.5	16.1	37.5	42.8
5	34	55	21	41	51.2	37.0	16.4	38.0	43.2
6	29	47	18	46	39.1	34.4	17.8	40.6	43.9
7	35	58	23	40	57.5	37.5	16.1	37.5	42.8
8	42	57	15	33	45.5	41.1	15.1	33.9	44.5
9	41	54	13	34	38.2	40.6	15.5	34.4	44.5
10	36	60	24	39	61.5	38.0	15.8	37.0	42.6
11	49	58	9	26	34.6	43.2	13.9	31.8	43.8
12	48	61	13	27	48.1	44.2	13.8	30.8	45.0
13	48	60	12	27	44.4	44.2	13.9	30.8	45.3

Students	Pretest score	Posttest score	Crude gain	Maximum possible gain	Crude gain ratio	Estmtd. true pretest	Estmtd. true gain	Estmtd. maximum true gain	Estmtd. true gain ratio
14	39	55	16	36	44.4	39.6	15.7	35.4	44.3
15	37	48	11	38	28.9	38.5	16.0	36.5	44.0
16	30	55	25	45	55.6	34.9	17.1	40.1	42.5
17	48	54	6	27	22.2	44.2	14.4	30.8	46.8
18	47	58	11	28	39.3	43.7	14.2	31.3	45.5
19	47	55	8	28	28.6	43.7	14.5	31.3	46.2
20	43	59	16	32	50.0	41.6	14.8	33.4	44.3
21	42	56	14	33	42.4	41.1	15.1	33.9	44.7
22	50	55	5	25	20.0	45.3	14.0	29.7	47.1
23	43	59	16	32	50.0	41.6	14.8	33.4	44.3
24	39	54	15	36	41.7	39.6	15.8	35.4	44.5
25	41	58	17	34	50.0	40.6	15.1	34.4	44.0
26	38	59	21	37	56.8	39.0	15.5	36.0	43.2

Students	Pretest score	Posttest score	Crude gain	Maximum possible gain	Crude gain ratio	Estmtd. true pretest	Estmtd. true gain	Estmtd. maximum true gain	Estmtd. true gain ratio
27	37	53	16	38	42.1	38.5	16.1	36.5	44.3
28	43	59	16	32	50.0	41.6	14.8	33.4	44.3
29	43	63	20	32	62.5	41.6	14.5	33.4	43.3
30	43	55	12	32	37.5	41.6	15.1	33.4	45.2
31	35	48	13	40	32.5	37.5	16.8	37.5	44.9
32	40	53	13	35	37.1	40.1	15.1	34.9	44.9
33	43	55	12	32	37.5	41.6	15.7	33.4	45.2
34	44	54	10	31	32.3	42.2	15.0	32.8	45.7
35	32	60	28	43	65.1	35.9	16.4	39.1	41.9
36	32	48	16	43	37.2	35.9	17.3	39.1	44.3
37	42	52	10	33	30.3	41.1	15.5	33.9	45.6
38	37	57	20	38	52.6	38.5	15.8	36.5	43.4
39	40	51	11	35	31.4	40.1	15.8	34.9	45.4

APPENDIX N

TEST ITEM-OBJECTIVE REFERENCE: ADMINISTRATIVE PRETEST

Students	Pretest score	Posttest score	Crude gain	Maximum possible gain	Crude gain ratio	Estmtd. true pretest	Estmtd. true gain	Estmtd. maximum true gain	Estmtd. true gain ratio
40	42	59	17	33	51.5	41.1	14.9	33.9	44.0
41	38	56	18	37	48.6	39.0	15.8	36.0	43.8
42	34	46	12	41	29.3	37.0	17.1	38.0	45.1
43	42	55	13	33	39.4	41.1	15.2	33.9	44.9
44	40	59	19	35	54.3	40.1	15.2	34.9	43.6
Mean	40.2	55.6	15.4	34.8	44.3	40.1	15.4	34.9	44.4

APPENDIX N

TEST ITEM OBJECTIVE REFERENCE (ADMINISTRATIVE PRETEST)

TEST ITEM ID	OBJECTIVE ID	PRETEST ITEM NO.	PRETEST ITEM ID	NO. CORRECT RESPONSES	PER CENT CORRECT
3001104	1000102	1	3200101	25	56.82
3002516	1002506	2	3200502	3	6.82
3001315	1001304	3	3201303	22	50.00
3001324	2021803	4	3201804	3	6.82
3003517	1003508	5	3203505	30	68.18
3001615	1001604	6	3201606	26	59.09
3005409	2025401	7	3205407	10	22.73
3003313	2003306	8	3203308	9	20.45
3001109	1001102	10	3201110	19	43.18
3004313	1004305	11	3204311	40	90.91
3002217	1002209	12	3202212	10	22.73
3005617	2045602	13	3205613	25	56.82
3002519	2042503	14	3202514	23	52.27
3004323	2074031	15	3204315	25	56.82
3001418	2061401	16	3201416	32	72.73
3002614	2052603	17	3202617	6	13.64
3005818	2035801	18	3205818	38	86.36
3002021	2062001	19	3202019	16	36.36
3002824	2102802	20	3202820	27	61.36
3006214	2026202	21	3206221	38	86.36
3004322	2064302	22	3204323	39	88.64
3005023	1005004	23	3205023	40	90.91
3002316	2072902	24	3202924	9	20.45
3003918	2040901	25	3209925	20	45.45
3002225	2012201	26	3202226	3	6.82
3003425	2060404	27	3204427	19	43.18
3002818	1002802	28	3201828	39	88.64
3003119	2032002	29	3203129	10	22.73
3004731	1004706	31	3204731	31	70.45
3003231	2032207	32	3203232	13	29.55
3003932	2043902	34	3203934	22	50.00
3005335	2015304	35	3205335	20	45.45
3004236	2074002	36	3204236	29	65.91
3003737	2033701	37	3203737	17	38.64
3004538	2126503	38	3204538	36	81.82
3002339	2042303	39	3202339	20	45.45
3005641	2035002	41	3205641	38	86.36
3001342	2013102	42	3201342	17	38.64

TEST ITEM ID	OBJECTIVE ID	PRETEST ITEM NO.	PRETEST ITEM ID	NO. CORRECT RESPONSES	PER CENT CORRECT
3004115	2034105	43	3204143	6	13.64
3003317	2033302	44	3203344	20	45.45
3005207	2025203	45	3205245	37	84.09
3001509	2051501	46	3201546	16	36.36
3003205	1003201	47	3203247	13	29.55
3004614	1004607	48	3204648	26	59.09
3003805	2023801	49	3203849	39	88.64
3000503	1000502	50	3200550	30	68.18
3001706	2011705	51	3201751	24	54.55
3003713	2053703	52	3203752	31	70.45
3002312	2032301	54	3202354	9	20.45
3003112	2023103	55	3203155	14	31.82
3003912	1003902	56	3203956	20	45.45
3005512	2035502	57	3205557	13	29.55
3004821	2084803	58	3204858	36	81.82
3004210	2034203	59	3204259	27	61.36
3002716	2042702	60	3202760	16	36.36
3001813	2021801	61	3201861	36	81.82
3004410	2034403	62	3204462	10	22.73
3000206	1000203	63	3200263	27	61.36
3002805	1002802	64	3202864	34	77.27
3005118	2075101	65	3205165	43	97.73
3000912	1000903	66	3200966	31	70.45
3003424	2043402	67	3203467	31	70.45
3001411	2041404	68	3201468	7	15.91
3005713	2025701	69	3205769	28	63.64
3004511	2064505	70	3204570	22	50.00
3000407	1000402	71	3200471	10	22.73
3002110	2032103	72	3202172	31	70.45
3005401	1005401	73	3205473	36	81.82
3000122	1000111	74	3200174	27	61.36
3001919	2041903	75	3201975	20	45.45
3001008	1001003	76	3201076	38	86.36
3003503	1003502	77	3203577	42	95.45
3005007	1005004	78	3205078	40	90.91
3001213	1001203	79	3201279	14	31.82
3003013	2033003	80	3203080	15	34.09

APPENDIX O

TEST ITEM-OBJECTIVE REFERENCE: ADMINISTRATIVE POSTTEST

APPENDIX O

TEST ITEM OBJECTIVE REFERENCE (ADMINISTRATIVE POSTTEST)

TEST ITEM ID	OBJECTIVE ID	PRETEST ITEM NO.	POSTTEST ITEM ID	NO. CORRECT RESPONSES	PER CENT CORRECT	TEST ITEM ID	OBJECTIVE ID	PRETEST ITEM NO.	POSTTEST ITEM ID	NO. CORRECT RESPONSES	PER CENT CORRECT
3000104	1000102	1	3300142	29	65.91	3004115	2034105	43	3304121	35	79.55
3000516	1000500	2	3300517	5	11.36	3000317	2030302	44	3300310	38	86.36
3001315	1001304	3	3301331	5	11.36	3005207	2025203	45	3305216	41	93.18
3001824	2031803	4	3301865	6	13.64	3001509	2051501	46	3301568	31	70.45
3003517	1003508	5	3303561	34	77.27	3003205	1003201	47	3303264	27	61.36
3001615	1001604	6	3301634	23	52.27	3004614	1004607	48	3304615	21	47.73
3005409	2025401	7	3305459	33	75.00	3003805	2023001	49	3303869	44	100.00
3003313	2063306	8	3303307	30	68.18	3000503	1000502	50	3300511	41	93.18
3001109	1001102	10	3301139	38	86.36	3001706	2011705	51	3301776	42	95.45
3004913	1004905	11	3304967	44	100.00	3003710	2053703	52	3303749	38	86.36
3000217	1000209	12	3300225	2	4.55	3002312	2032301	54	3302306	17	38.64
3005617	2045602	13	3305678	31	70.45	3003112	2023103	55	3303180	17	38.64
3002519	2042503	14	3302503	44	100.00	3003912	1003902	56	3303947	24	54.55
3004523	2074501	15	3304541	33	75.00	3005512	2035502	57	3305535	43	97.73
3001418	2061401	16	3301430	41	93.18	3004821	2084803	58	3304845	43	97.73
3002614	2052601	17	3302627	43	97.73	3004210	2034203	59	3304226	37	84.09
3005818	2035801	18	3305836	31	70.45	3002716	2042702	60	3302775	30	68.18
3002021	2062001	19	3302008	43	97.73	3001813	2021801	61	3301874	42	95.45
3002824	2102802	20	3302851	20	45.45	3004410	2034403	62	3304458	38	86.36
3000610	2020602	21	3300643	41	93.18	3000206	1000203	63	3300237	36	81.82
3004322	2064302	22	3304324	28	63.64	3002805	1002802	64	3302823	41	93.18
3005112	1005006	23	3305063	43	97.73	3005118	2075101	65	3305129	44	100.00
3002516	2072902	24	3302912	31	70.45	3000912	1000903	66	3300960	34	77.27
3000918	2040901	25	3300948	30	68.18	3003424	2063402	67	3303422	42	95.45
3002205	2012201	26	3302244	33	75.00	3001411	2041404	68	3301438	17	38.64
3001425	2060404	27	3300413	33	75.00	3005713	2025701	69	3305740	43	97.73
3000808	1000802	28	3300861	41	93.18	3004511	2064505	70	3304566	29	65.91
3003119	2053702	29	3303073	24	54.55	3000407	1000402	71	3300471	14	31.82
3004721	1004705	31	3304704	44	100.00	3002110	2032103	72	3302172	39	88.64
3003227	2053207	32	3303262	19	43.18	3005401	1005401	73	3305456	43	97.73
3003922	2043902	34	3303957	33	75.00	3000122	1000111	74	3300133	22	50.00
3005320	2015204	35	3305320	40	90.91	3001919	2041903	75	3301954	24	54.55
3004219	2074202	36	3304277	38	86.36	3001008	1001003	76	3301053	43	97.73
3003726	2133701	37	3303709	41	93.18	3003503	1003502	77	3303519	44	100.00
3004922	2124903	38	3304552	38	86.36	3005007	1005004	78	3305014	41	93.18
3002319	2042302	39	3302318	26	59.09	3001210	1001203	79	3301246	34	77.27
3005922	2035902	41	3305970	42	95.45	3003010	2033003	80	3303028	27	61.36
3001313	2031301	42	3301355	22	50.00						

APPENDIX P

EXAMPLE: STUDENT TRACK

APPENDIX P

STUDENT TRACK REPORT

NIXON L R STUDENT ID=716396 RESEARCH ID= 3

PART	SEGMENT	MODULE	NO. ITEMS		ITERATION 1 RS	ITERATION 1 PC	ITERATION 2 RS	ITERATION 2 PC
1	1	1	10		8	80.0	0	0.0
1	2	1	9		3	33.3	0	00.0
		**********	19	PART TOTALS	11		0	
2	1	1	10		9	90.0	0	0.0
2	2	1	10		7	70.0	6	60.0
2	3	1	10		5	50.0	6	60.0
2	4	1	10		8	80.0	0	0.0
2	5	1	10		5	50.0	9	90.0
2	6	1	10		8	80.0	0	0.0
2	7	1	10		9	90.0	0	0.0
2	8	1	10		10	100.0	0	0.0
2	9	1	10		4	40.0	10	100.0
		**********	90	PART TOTALS	63		31	
3	1	2	10		8	80.0	0	0.0
3	2	2	7		5	71.4	0	0.0
3	3	2	10		9	90.0	0	0.0
3	4	2	10		9	90.0	0	0.0
3	5	1	10		8	80.0	0	0.0

APPENDIX Q

TOTAL % OBJECTIVES ACHIEVED BY EACH STUDENT BEFORE AND AFTER REMEDIATION

APPENDIX Q

TOTAL % OBJECTIVES ACHIEVED BY EACH STUDENT BEFORE AND AFTER REMEDIATION

Student	% Objectives before remediation	% Objectives after remediation
1	80.6	85.1
2	78.5	84.0
3	77.9	83.4
4	80.4	86.0
5	74.0	81.3
6	70.6	76.2
7	76.4	83.4
8	75.3	81.5
9	82.5	85.3
10	75.1	83.2
11	78.9	87.5
12	78.3	84.3
13	84.0	90.8
14	78.3	85.1
15	77.4	82.1
16	75.5	86.8
17	79.8	85.8
18	80.4	85.8
19	66.8	77.0

Student	% Objectives before remediation	% Objectives after remediation
20	77.2	82.3
21	73.4	80.2
22	80.4	86.4
23	81.9	85.5
24	84.0	87.0
25	73.6	80.4
26	82.1	85.3
27	73.8	79.8
28	81.5	84.5
29	81.7	86.8
30	80.4	87.9
31	74.7	79.4
32	81.5	86.6
33	80.0	85.8
34	79.1	84.9
35	73.6	81.7
36	77.7	84.3
37	77.2	81.9
38	73.8	82.8

Student	% Objectives before remediation	% Objectives after remediation
39	79.1	80.9
40	82.8	85.8
41	70.4	81.5
42	72.6	80.4
43	80.2	85.5
44	74.5	82.8
Mean %	77.7	83.8

APPENDIX R

EXAMPLE: TEST ITEM-OBJECTIVE REFERENCE: PROGRESS CHECK

APPENDIX R

TEST ITEM OBJECTIVE REFERENCE--PROGRESS CHECK

PART- 1, SEGMENT- 2

EO NO. / TO NO.	OBJECTIVE IDENTIFIER	TEST ITEM IDENTIFIER	--- ITERATION 1 --- NO. CORRECT RESPONSES	--- ITERATION 1 --- PERCENT CORRECT	ITERATION 2 NO. CORRECT RESPONSES	TOTAL NO. CORRECT	TOTAL PC CORRECT
TO-01	1000201	3400201	41	93.2	2	43	97.7
TO-02	1000202	3400202	25	56.8	11	36	81.8
EO-01	2040201						
TO-04	1000204	3400203	41	93.2	3	44	100.0
TO-05	1000205	3400204	39	88.6	5	44	100.0
EO-01	2060201						
TO-06	1000206	3400205	6	13.6	26	32	72.7
EO-01	2080201						
TO-08	1000208	3400206	10	22.7	18	28	63.6
TO-11	1000211	3400208	35	79.5	9	44	100.0
EO-01	2120201						
TO-12	1000212	3400209	34	77.3	8	42	95.5
TO-13	1000213	3400210	32	72.7	7	39	88.6

N1 = 44 N2 = 37

APPENDIX S

MODULE QUESTIONNAIRE

APPENDIX S

MODULE QUESTIONNAIRE

PART_____
SEGMENT_____
MODULE ____

NAME _____
ID# _____
DATE _____

Exact Time Spent on Instructional Material in Module: ____Minutes.

Answer the starred items only if applicable.
Complete all but Question 9 after working through instructional material.
Turn in questionnaire after completing the Progress Check.

From your own point of view:	High	Above Avg	Avg	Below Avg	Low
1. Was the material interesting?					
2. Was the material difficult?					
3. Were the questions difficult?					
*4. Was the video interesting?					
*5. Was the video lecturer interesting?					
*6. Was the audio presentation interesting?					
*7. Was the audio lecturer interesting?					
*8. Was the programed sequence interesting?					
9. Was the progress check difficult?					

10. If possible, specify where you had difficulty.

11. Comments and Suggestions.

APPENDIX T

PART AND SEGMENT: AVERAGE % PROGRESS CHECK
ITEMS CORRECT, STUDENT TIME, AND % STUDENT INTEREST

APPENDIX T

PART AND SEGMENT: AVERAGE % PROGRESS CHECK ITEMS CORRECT, STUDENT TIME, AND % STUDENT INTEREST

Part and segment number	Number of test items	Number of students	Mean raw score	Average % PC items correct	Average student time (min.)	% Student interest (high)	% Student interest (low)
1.1	10	44	8.2	82.0	91.5	73.3	6.7
1.2	9	44	6.0	66.7	30.3	39.5	10.5
Unweighted mean % for Part One				74.4	60.9	56.4	8.6
2.1	10	44	7.8	78.0	37.4	65.8	9.8
2.2	10	44	6.8	68.0	31.4	32.6	16.3
2.3	10	44	6.5	65.0	31.3	40.0	13.3
2.4	10	43	7.1	71.0	25.4	55.6	11.1
2.5	10	44	6.5	65.0	26.7	40.0	20.0
2.6	10	44	7.1	71.0	34.7	68.2	2.3
2.7	10	44	6.3	63.0	29.7	62.8	0.0
2.8	10	44	8.9	89.0	34.0	80.9	2.4
2.9	10	44	6.5	65.0	54.8	48.5	24.2

Part and segment number	Number of test items	Number of students	Mean raw score	Average % PC items correct	Average student time (min.)	% Student interest (high)	% Student interest (low)
Unweighted mean % for Part Two				70.6	33.9	54.9	11.0
3.1	10	44	6.6	66.0	36.6	27.5	22.5
3.2	7	44	4.6	65.7	24.8	31.7	24.4
3.3	10	44	7.5	75.0	31.6	31.7	19.5
3.4	10	43	7.9	79.0	26.8	27.9	23.3
3.5	10	43	8.4	84.0	34.4	61.5	7.7
Unweighted mean % for Part Three				73.9	30.8	36.0	19.4
4.1	9	44	7.5	83.3	38.5	50.0	21.5
4.2	10	44	8.5	85.0	43.2	42.9	16.7
4.3	10	44	8.7	87.0	41.0	27.5	15.0
4.4	10	44	9.2	92.0	55.5	33.3	26.2
4.5	9	44	7.4	82.2	49.5	41.4	7.3
4.6	10	44	7.3	73.0	41.1	60.9	4.9
4.7	10	44	7.7	77.0	42.8	56.4	7.7

Part and segment number	Number of test items	Number of students	Mean raw score	Average % PC items correct	Average student time (min.)	% Student interest (high)	% Student interest (low)
Unweighted mean % for Part Four				82.8	44.5	44.6	14.1
5.1	9	44	7.0	77.8	57.4	25.6	25.6
5.2	8	44	5.6	70.0	58.6	41.8	23.3
5.3	10	44	7.4	74.0	39.3	28.2	12.9
5.4	11	44	7.6	69.1	50.5	17.9	38.5
5.5	9	44	5.3	58.9	52.0	4.7	37.2
5.6	8	44	5.6	70.0	40.3	25.0	22.2
5.7	8	44	6.4	80.0	31.0	21.4	9.6
5.8	9	44	8.1	90.0	38.9	23.8	16.7
5.9	9	43	7.0	77.8	55.4	26.2	16.7
5.10	7	43	4.6	65.7	35.3	9.5	35.7

Part and segment number	Number of test items	Number of students	Mean raw score	Average % PC items correct	Average student time (min.)	% Student interest (high)	% Student interest (low)
Unweighted mean % for Part Five				73.3	45.8	22.4	23.8
6.1	14	44	11.2	80.0	48.0	63.4	4.9
6.2	6	44	4.9	81.7	46.6	44.2	4.7
6.3	12	44	10.9	90.8	45.2	32.6	11.6
6.4	8	44	7.0	87.5	40.8	27.5	2.5
Unweighted mean % for Part Six				85.0	45.1	41.9	5.9
7.1	8	44	5.5	68.8	35.7	15.9	34.0
7.2	4	44	3.3	82.5	42.0	34.2	26.9
7.3	9	44	6.9	76.7	38.0	36.6	2.4
7.4	6	44	4.7	78.3	42.1	22.7	34.0
7.5	7	44	5.2	74.3	38.8	39.5	9.3

Part and segment number	Number of test items	Number of students	Mean raw score	Average % PC items correct	Average student time (min.)	% Student interest (high)	% Student interest (low)
Unweighted mean % for Part Seven				76.1	39.3	29.7	21.3
8.1	10	44	8.2	82.0	58.8	38.1	16.7
8.2	13	42	10.5	80.8	60.2	31.7	14.6
8.3	8	44	7.4	92.5	52.1	51.3	7.7
8.4	8	44	6.1	76.3	63.0	20.0	20.0
8.5	11	43	8.5	77.3	54.9	16.7	22.3
8.6	6	44	4.6	76.7	55.5	36.1	19.5
Unweighted mean % for Part Eight				80.9	57.4	32.3	16.8
9.1	5	44	4.0	80.0	41.6	42.9	2.4
9.2	5	44	4.4	88.0	40.0	32.5	7.0

Part and segment number	Number of test items	Number of students	Mean raw score	Average % PC items correct	Average student time (min.)	% Student interest (high)	% Student interest (low)
Unweighted mean % for Part Nine				84.0	40.8	37.7	4.7
10.1	7	44	6.5	92.9	51.7	40.0	10.0
10.2	10	44	8.9	89.0	45.4	35.9	5.1
Unweighted mean % for Part Ten				91.0	48.5	37.9	7.5
11.1	10	44	8.0	80.0	30.8	30.8	0.0
11.2	8	44	7.3	91.3	29.8	23.1	10.3
11.3	4	44	3.3	82.5	33.9	40.5	8.1
Unweighted mean % for Part Eleven				84.6	31.5	31.4	6.1
12.1	10	44	8.7	87.0	68.4	70.7	2.4
12.2	9	44	7.7	85.6	70.9	55.5	6.7
12.3	10	43	9.4	94.0	37.7	64.3	0.0
12.4	10	44	8.5	85.0	40.6	53.6	4.9

Part and segment number	Number of test items	Number of students	Mean raw score	Average % PC items correct	Average student time (min.)	% Student interest (high)	% Student interest (low)
Unweighted mean % for Part Twelve				87.9	54.4	61.0	3.5
Unweighted mean % for all parts				78.4	43.5	39.3	14.2

APPENDIX U

AVERAGE PERCENT PROGRESS CHECK ITEMS CORRECT,
STUDENT TIME AND % STUDENT INTEREST BY SEGMENT WITHIN MEDIA

APPENDIX U

AVERAGE % PROGRESS CHECK ITEMS CORRECT, STUDENT TIME, AND % STUDENT INTEREST BY SEGMENT WITHIN MEDIA

AUDIOTAPE

Part and segment number	Number of test items	Number of students	Mean raw score	Average % PC items correct	Average student time (min.)	% Student interest (high)	% Student interest (low)
2.2	10	21	7.25	72.5	31.4	32.5	17.5
2.3	10	19	6.35	63.5	31.3	53.5	6.3
2.4	10	20	7.30	73.0	25.4	70.8	11.1
2.5	10	21	6.90	69.0	26.7	42.3	15.0
3.1	10	21	6.55	65.5	36.6	19.3	22.2
3.2	7	21	5.00	71.5	24.8	43.2	23.2
3.3	10	21	7.60	76.0	31.6	40.0	21.3
3.4	10	21	8.10	81.0	26.8	20.0	36.8
5.7	8	22	6.60	82.5	31.0	14.7	10.1
5.8	9	22	8.04	89.4	38.9	14.6	14.6
5.9	9	22	6.95	77.2	55.4	31.9	22.8

AUDIOTAPE

Part and segment number	Number of test items	Number of students	Mean raw score	Average % PC items correct	Average student time (min.)	% Student interest (high)	% Student interest (low)
5.10	7	22	4.70	67.2	35.3	9.1	36.3
7.1	8	23	5.40	67.5	68.8	13.3	35.3
7.2	4	23	3.20	80.0	82.5	23.8	29.3
7.3	5	23	4.05	81.1	76.7	32.8	5.0
7.4	6	23	4.75	79.2	78.3	29.9	34.5
Unweighted mean % for R segments				74.8	43.8	30.7	21.3
Unweighted mean % for all segments				74.8	43.8	30.7	21.3

VIDEOTAPE

Part and segment number	Number of test items	Number of students	Mean raw score	Average % PC items correct	Average student time (min.)	% Student interest (high)	% Student interest (low)
2.2	10	23	6.50	65.0	31.4	30.4	19.9
2.3	10	23	6.75	67.5	31.3	39.3	20.2
2.4	10	23	7.00	70.0	25.4	38.2	14.1
2.5	10	23	6.20	62.0	26.7	39.4	21.6
3.1	10	23	6.65	66.5	36.6	33.2	23.2
3.2	7	23	4.30	61.5	24.8	20.9	22.8
3.3	10	23	7.40	74.0	31.6	25.8	17.8
3.4	10	22	7.80	78.0	26.8	36.7	9.2
5.7	8	22	6.15	72.0	31.0	27.3	9.1
5.8	9	22	8.05	89.5	38.9	34.1	18.7
5.9	9	21	7.78	77.8	55.4	20.0	10.0

VIDEOTAPE

Part and segment number	Number of test items	Number of students	Mean raw score	Average % PC items correct	Average student time (min.)	% Student interest (high)	% Student interest (low)
5.10	7	22	7.00	62.9	35.3	10.0	35.0
7.1	8	22	5.70	71.3	68.8	19.1	32.8
7.2	4	22	3.55	88.8	82.5	47.2	22.3
7.3	9	22	6.40	71.1	76.7	40.0	0.0
7.4	6	22	4.55	75.9	78.3	14.6	32.8
Unweighted mean % for R segments				72.1	43.8	29.8	19.3
7.5	7	44	5.2	74.3	38.8	39.5	9.3
9.1	5	44	4.0	80.0	41.6	42.9	2.4
9.2	5	44	4.4	88.0	40.0	32.5	7.0
Unweighted mean % for NR segments				80.8	40.1	38.3	6.2
Unweighted mean % for all segments				73.5	43.2	31.1	17.3

AUDIOTAPE IP

Part and segment number	Number of test items	Number of students	Mean raw score	Average % PC items correct	Average student time (min.)	% Student interest (high)	% Student interest (low)
4.4	10	44	9.2	92.0	55.5	33.3	26.2
4.5	9	44	7.4	82.2	49.5	41.4	7.3
4.6	10	44	7.3	73.0	41.1	60.9	4.9
4.7	10	44	7.7	77.0	42.8	56.4	7.7
Unweighted mean % for R segments				81.1	47.2	48.0	11.5
10.1	7	44	6.5	92.9	51.7	40.0	10.0
10.2	10	44	8.9	89.0	45.4	35.9	5.1
Unweighted mean % for NR segments				90.9	48.6	37.9	7.5
Unweighted mean % for all segments				84.4	47.7	44.7	10.2

CAI

Part and segment number	Number of test items	Number of students	Mean raw score	Average % PC items correct	Average student time (min.)	% Student interest (high)	% Student interest (low)
12.1	10	44	8.7	87.0	68.4	70.7	2.4
12.2	9	44	7.7	85.6	70.9	55.5	6.7
12.3	10	43	9.4	94.0	37.7	64.3	0.0
12.4	10	44	8.5	85.0	40.6	53.6	4.9
Unweighted mean % for R segments				87.9	54.4	61.0	3.5
Unweighted mean % for all segments				87.9	54.4	61.0	3.5

LAS

Part and segment number	Number of test items	Number of students	Mean raw score	Average % PC items correct	Average student time (min.)	% Student interest (high)	% Student interest (low)
8.4	8	44	6.1	76.3	63.0	20.0	20.0
8.5	11	43	8.5	77.3	54.9	16.7	22.3
8.6	6	44	4.6	76.7	55.5	36.1	19.5
Unweighted mean % for R segments				76.8	57.8	24.3	20.6
2.9	10	44	6.5	65.0	54.8	48.5	24.2
Unweighted mean % for NR segments				65.0	54.8	48.5	24.2
Unweighted mean % for all segments				73.8	57.1	30.3	21.5

LINEAR TEXT

Part and segment number	Number of test items	Number of students	Mean raw score	Average % PC items correct	Average student time (min.)	% Student interest (high)	% Student interest (low)
4.1	9	44	7.5	83.3	38.5	50.0	21.5
4.2	10	44	8.5	85.0	43.2	42.9	16.7
4.3	10	44	8.7	87.0	41.0	27.5	15.0
5.4	11	44	7.6	69.1	50.5	17.9	38.5
5.5	9	44	5.3	58.9	52.0	4.7	37.2
5.6	8	44	5.6	70.0	40.3	25.0	22.2
8.1	10	44	8.2	82.0	58.8	38.1	16.7
8.2	13	42	10.5	80.8	60.2	31.7	14.6
8.3	8	44	7.4	92.5	52.1	51.3	7.7
Unweighted mean % for R segments				78.7	48.5	32.1	21.1
Unweighted mean % for all segments				78.7	48.5	32.1	21.1

SYNDACTIC TEXT

Part and segment number	Number of test items	Number of students	Mean raw score	Average % PC items correct	Average student time (min.)	% Student interest (high)	% Student interest (low)
2.6	10	44	7.1	71.0	34.7	68.2	2.3
2.7	10	44	6.3	63.0	29.7	62.8	0.0
2.8	10	44	8.9	89.0	34.0	80.9	2.4
6.1	14	44	11.2	80.0	48.0	63.4	4.9
6.2	6	44	4.9	81.7	46.6	44.2	4.7
6.3	12	44	10.9	90.8	45.2	32.6	11.6
11.1	10	44	8.0	80.0	30.8	30.8	0.0
11.2	8	44	7.3	91.3	29.8	23.1	10.3
11.3	4	44	3.3	82.5	33.9	40.5	8.1
Unweighted mean % for R segments				81.0	36.9	49.6	4.9

SYNDACTIC TEXT

Part and segment number	Number of test items	Number of students	Mean raw score	Average % PC items correct	Average student time (min.)	% Student interest (high)	% Student interest (low)
1.1	10	44	8.2	82.0	91.5	73.3	6.7
2.1	10	44	7.8	78.0	37.4	65.8	9.8
3.5	10	43	8.4	84.0	34.4	61.5	7.7
5.1	9	44	7.0	77.8	57.4	25.6	25.6
5.2	8	44	5.6	70.0	58.6	41.8	23.3
5.3	10	44	7.4	74.0	39.3	28.2	12.9
5.4	8	44	7.0	87.5	40.8	17.9	38.5
Unweighted mean % for NR segments				79.0	51.3	44.9	17.8
Unweighted mean % for all segments				80.2	43.2	47.5	10.6

APPENDIX V

PERCENT STUDENTS SCORING 70% OR 80% ON PROGRESS CHECKS BEFORE AND AFTER REMEDIATION BY SEGMENT WITHIN MEDIA

PERCENTAGE OF STUDENTS SCORING 70% or 80% ON PROGRESS CHECKS BEFORE AND AFTER REMEDIATION BY SEGMENT WITHIN MEDIA

AUDIOTAPE-VIDEOTAPE

Part and segment number	Number of test items	Number of students	% Students 70% or better before remed.	% Students 80% or better before remed.	% Students 80% or better after remed.
2.2	10	44	63.6	31.8	70.5
2.3	10	44	56.8	18.2	56.8
2.4	10	43	72.1	34.9	72.1
2.5	10	44	52.3	25.0	63.6
3.1	10	44	54.5	22.7	45.5
3.2	7	44	61.4	20.5	38.6
3.3	10	44	77.3	63.6	84.1
3.4	10	43	93.0	72.1	95.3
5.7	8	44	79.5	52.3	77.3
5.8	9	44	95.5	77.3	79.5
5.9	9	43	67.4	39.5	76.7
5.10	7	43	58.1	20.9	51.2
7.1	8	44	59.1	18.2	43.2
7.2	4	44	88.6	45.5	75.0
7.3	9	44	59.1	34.1	45.5
7.4	6	44	59.1	59.1	79.5
7.5	7	44	79.5	43.2	70.5
9.1	5	44	68.2	68.2	79.5
9.2	5	44	93.2	93.2	100.0
Unweighted mean % students			69.9	44.3	63.9

AUDIOTAPE-IP BOOKLET

Part and segment number	Number of test items	Number of students	% Students 70% or better before remed.	% Students 80% or better before remed.	% Students 80% or better after remed.
4.4	10	44	97.7	95.5	97.7
4.5	9	44	81.8	43.2	54.5
4.6	10	44	72.7	52.3	79.5
4.7	10	44	84.1	68.2	93.2
10.1	7	44	100.0	90.9	100.0
10.2	10	44	100.0	95.5	100.0
Unweighted mean % students			89.4	74.2	87.5

CAI

Part and segment number	Number of test items	Number of students	% Students 70% or better before remed.	% Students 80% or better before remed.	% Students 80% or better after remed.
12.1	10	44	95.5	84.1	93.2
12.2	9	44	93.2	59.1	70.5
12.3	10	43	100.0	97.7	100.0
12.4	10	44	90.9	84.1	95.5
Unweighted mean % students			94.9	81.2	89.8

LAS

Part and segment number	Number of test items	Number of students	% Students 70% or better before remed.	% Students 80% or better before remed.	% Students 80% or better after remed.
2.9	10	44	47.7	36.4	79.5
8.4	8	44	65.9	43.2	79.5
8.5	11	43	76.7	60.5	93.0
8.6	6	44	56.8	56.8	88.6
Unweighted mean % students			61.7	49.2	85.1

LINEAR TEXT

Part and segment number	Number of test items	Number of students	% Students 70% or better before remed.	% Students 80% or better before remed.	% Students 80% or better after remed.
4.1	9	44	86.4	61.4	72.7
4.2	10	44	97.7	88.6	95.5
4.3	10	44	90.9	88.6	95.5
5.4	11	44	54.5	25.0	47.7
5.5	9	44	9.1	4.5	11.4
5.6	8	44	65.9	13.6	40.9
8.1	10	44	84.1	70.5	84.1
8.2	13	42	78.6	42.9	66.7
8.3	8	44	100.0	88.6	97.7
Unweighted mean % students			74.1	53.7	68.0

SYNDACTIC TEXT

Part and segment number	Number of test items	Number of students	% Students 70% or better before remed.	% Students 80% or better before remed.	% Students 80% or better after remed.
1.1	10	44	90.9	70.5	93.2
2.1	10	44	79.5	65.9	97.7
2.6	10	44	68.2	40.9	72.7
2.7	10	44	36.4	11.4	43.2
2.8	10	41	97.7	86.4	93.2
3.5	10	43	95.3	83.7	100.0
5.1	9	44	68.2	43.2	81.8
5.2	8	44	54.5	22.7	70.5
5.3	10	44	75.0	52.3	86.4
6.1	14	44	86.4	50.0	75.0
6.2	6	44	72.7	72.7	90.9
6.3	12	44	100.0	95.5	100.0
6.4	8	44	95.5	79.5	95.5
11.1	10	44	84.1	70.5	84.1
11.2	8	44	100.0	88.6	90.9
11.3	4	44	86.4	52.3	70.5
Unweighted mean % students			80.7	61.7	84.1

APPENDIX W

EXAMPLE: SYNDACTIC TEXT FRAME ANALYSIS

APPENDIX W

SYNDACTIC TEXT FRAME ANALYSIS DATA

NO. M'n Taking this version 13

PART 2 SEGMENT 6

☐ Single volume

☒ 4-Volume

SUMMARY PRE-QUIZ 1

1	2
2	5
3	7
4	3
5	4
6	⊘

PROGRAMMED SEQUENCE 1
Taken by 6 M'n.

1	0	11	3	21	5	31		41		51	
2	2	12	6	22	0	32		42		52	
3	4	13	0	23	6	33		43			
4	1	14	0	24		34		44			
5	0	15	0	25		35		45			
6	0	16	3	26		36		46			
7	2	17	0	27		37		47			
8	0	18	4	28		38		48			
9	5	19	0	29		39		49			
10	0	20	0	30		40		50			

SUMMARY POST-QUIZ 1

3	0
5	0
4	1
2	1
1	0
	⊘

SUMMARY PRE-QUIZ 2

1	4
2	0
3	0
4	2
5	3
6	⊘

PROGRAMMED SEQUENCE 2
Taken by 5 M'n.

1	0	11	1	21		31		41	
2	0	12	0	22		32		42	
3	0	13	1	23		33		43	
4	0	14	0	24		34		44	
5	0	15	2	25		35		45	
6	0	16	2	26		36		46	
7	1	17	4	27		37			
8	2	18		28		38			
9	2	19		29		39			
10	2	20		30		40			

SUMMARY POST-QUIZ 2

3	2
1	0
5	0
4	0
2	0
	⊘

SUMMARY PRE-QUIZ 3

1	11
2	3
3	5
4	6
5	2
6	⊘

PROGRAMMED SEQUENCE 3
Taken by 2 M'n.

1	0	11	0	21	4	31	1	41	
2	0	12	0	22	1	32		42	
3	2	13	0	23	7	33		43	
4	1	14	4	24	3	34		44	
5	1	15	3	25	0	35		45	
6	1	16	1	26	3	36		46	
7	3	17	1	27	0	37		47	
8	0	18	3	28	2	38		48	
9	2	19	3	29	1	39		49	
10	1	20	0	30	3	40		50	

SUMMARY POST-QUIZ 3

2	3
4	1
5	1
1	0
3	0
	⊘

SUMMARY PRE-QUIZ 4

1	8
2	6
3	7
4	3
5	1
6	1

PROGRAMMED SEQUENCE 4
Taken by 8 M'n.

1	1	11	6	21		31		41		51	
2	2	12	1	22		32		42		52	
3	8	13	0	23		33		43		53	
4	0	14	1	24		34		44		54	
5	1	15	0	25		35		45		55	
6	0	16	0	26		36		46			
7	0	17	1	27		37		47			
8	2	18	2	28		38		48			
9	0	19	0	29		39		49			
10	1	20	0	30		40		50			

SUMMARY POST-QUIZ 4

3	5
4	5
2	1
1	2
6	0
5	0

SUMMARY PRE-QUIZ 5

1	
2	
3	
4	
5	
6	1

PROGRAMMED SEQUENCE 5
Taken by ___ M'n.

1		11		21	
2		12		22	
3		13		23	
4		14		24	
5		15		25	
6		16			
7		17			
8		18			
9		19			
10		20			

SUMMARY POST-QUIZ 5